PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : G01N 33/68 — A2

(11) International Publication Number: WO 00/68695

(43) International Publication Date: 16 November 2000 (16.11.00)

(21) International Application Number: PCT/US00/13246

(22) International Filing Date: 12 May 2000 (12.05.00)

(30) Priority Data:
60/134,017 12 May 1999 (12.05.99) US

(71) Applicant: THE BOARD OF REGENTS FOR OKLAHOMA STATE UNIVERSITY [US/US]; Oklahoma State University, 203 Whitehurst, Stillwater, OK 74078 (US).

(72) Inventors: PURDIE, Neil; 909 Greystone, Stillwater, OK 74074 (US). PROVINCE, Dennis, William; 820 Sunset Drive, Edmond, OK 73003 (US).

(74) Agent: WEEKS, R., Alan; Fellers, Snider, Blankenship, Bailey & Tippens, P.C., Suite 800, 321 S. Boston Avenue, Tulsa, OK 74103-3318 (US).

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: REAGENT AND PROCESS FOR PEPTIDE/PROTEIN DIFFERENTIATION USING CIRCULAR DICHROISM DETECTION

(57) **Abstract**

A reagent comprised of an aqueous solution of Cu(II)-D-histidine complex acts effectively as a devitalizing agent required to qualitatively identify an enantiomer and quantitatively determine its enantiomeric purities. The initial function of the host ligand (D-histidine) is to keep the Cu(II) ion in solution in high ph values. The base line CD spectrum associated with each Cu-(chiral ligand) host complex is uniquely different. On adding peptide or protein, exchange occurs between the host ligand (D-histidine) in the analyte ligand (protein). Exchanges produce changes in the CD spectrum that are significant enough that they have the potential of becoming a reliable spectroscopic fingerprint for every individual analyte.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

REAGENT AND PROCESS FOR PEPTIDE/PROTEIN DIFFERENTIATION USING CIRCULAR DICHROISM DETECTION

BACKGROUND OF THE INVENTION

Background of the Invention:

Fundamental to the structure of all peptides and proteins is the repeating peptide bond (-CO-NH_2) moiety. Molecules are produced by condensation of aminoacids that are naturally optically active. The addition of each aminoacid is accompanied by the introduction of another chiral center. For each new chiral central there are potentially two new stereoisomers. The peptide bond is a strong absorber of ultraviolet light. The UV spectral range, however, is inconvenient for analytical detection because of the wide range of interferences from other materials that absorb radiation in the same spectral range.

Historically, the most effective method developed fro the measurement of total protein in a clinical context involved binding the protein to Cu(II) ion present in a solution as the metal tartrate complex in a strongly basic aqueous medium (pH>12). This is conventionally known as the biuret reagent. The absorbance intensity of the purple color that is produced in the visible range of the spectrum by the metal-protein complex, correlates linearly with the amount of the protein present. The biuret test is still available commercially and used fro proteins assays. One great problem, however, is that absorbance detection is totally incapable of discriminating between enantiomers of the same chiral substance because their absorbance intensities are equivalent. The test can only be applied if the enantimers are first separated chromatographically.

A major new frontier in the pharmaceutical industry is the focus on the therapeutic properties of peptides and proteins as drugs. An innovative approach to drug

design as it exists in biotechnology is to substitute the unnatural D-enantimer of an aminoacid for the corresponding antrually occurring L-enantimer. It now becomes critically important to prove the stereochemistry in regulatory quality control (QC) laboratories.

Most known peptide analytical methods employ spectrophotometry. Spectrophotometry refers to the measurement of the absorption or transmission of incident light through solutions of test compounds. Typically, compounds of interest have characteristic spectra, transmitting or absorbing specific wave elengths of light, which can be used to determine the presence of these compounds or measure their concentration in test samples. Instruments designed for spectrophotometric absorption have a light source, for which the emitted wavelengths of light, which can be used to determine the presence of these compounds or measure their concentration in test samples. Instruments designed for spectrophotometric absorption have a light source, for which the emitted wavelength is known and may be adjusted, and one or more detectors sensitive to desired wavelengths of transmitted or reflected light. Spectrophotometric absorption can be used to determine the amount of a given compound that is present in a test sample.

Circular dichroism (CD) is a special type of absorption method in which the molecular composition of an analyte results in differential absorption of incident light not only at a specific wavelength but also of a particular polarization state. Circular dichroism is a chiroptical method which allows one to differentiate between different enantiomers; that is, optical isomers having one or more asymmetric carbon atom (chrial) centers. When utilizing CD, generally a sample is illuminated by two circularly polarized beams of light traveling in unison. Both beams pass through the sample simultaneously and are absorbed. If the sample is optically active, the beams are absorbed to different extents. The differences in absorption of the beams can then be displayed as a function of the wavelength of the incident light beam as a CD spectrum. No difference in absorption is observed fro optically inactive absorbers so that these compounds are not detected by a CD detecting system. The use of CD as a chiroptical method has been fully described in scientific literature, such as Lambert, J.B. et al. "Organic Structural

Analysis", Macmillan, New York, N.Y. 1976.

Early applications of the CD method primarily dealt with elucidation of molecular structures, especially natural products for which a technique capable of confirming or establishing absolute stereochemistry was critical. However, CD has also reportedly been used in a clinical method to quantitatively determine unconjugated bilirubin in blood plasma, Grahnen, A. et al. *Clinica Chimica Acta*, 52, 187-196 (1974). In the method thus disclosed, a complex was formed between bilirubin and human serum albumin as a CD probe for bilirubin analysis.

Clinical applications of circular dichroism are also discussed by Neil Purdie and Kathy A. Swallows in *Analytical Chemistry*, Vol. 61, No. 2, pp. 77A-89A (1989), herein incorporated by reference. However, suitable chemical reagents for carrying out such testing are not disclosed.

Modern Pharmaceutical and biotechnology conglomerates are committed to the production of chiral drug substances. Manufacturers have the option to prepare chiral drugs either as pure single substances (enantiomers) or as racemic mixtures. While racemates are ostensibly easier to produce, many good reasons exist for choosing to manufacture enantiomerically pure forms, not the least of which are considerations of the relative therapeutic and toxicity levels of each enantiomer either by itself or as half of a racemate. All of this means that the need exists for simple routine analytical methods that are adaptable to all chiral drug forms and used for regulatory control of their chemical and enantiomeric purities (EP), whether it is done by the manufacturer or by a Federal Agency.

Current analytical options that are applied to these tests generally involve simultaneous derivatizations of both enantiomers in a racemic or partial racemic mixture to their corresponding diastereoisomers by selective reactions with a third chiral species. Unlike enantiomers, diastereoisomers can be differentiated by physical properties other than just the direction of rotation of linearly polarized light. Chiral chromatography is a major player in the development of these methods. The chiral third party is introduced either in the mobile phase or immobilized on the stationary phase . Since the diastereomers elute after different retention times, achiral detectors, e.g. absorbance,

electrochemical and mass spectrometry, are sufficient to effect quantitative distinctions, within the limits of detection capabilities of the chosen methods.

If a preference exists to determine chemical and enantiomeric purities without a prior separation step, procedures generally call for the use of two detectors, one of which is a chiral detector such as polarimetry or circular dichroism (CD) . Using a combination of CD and absorbance detection, spectral differences or spectral ratios are among the strategies that can be used to manipulate data. Generally speaking most of these methods are based on single wavelength detection data.

By combining multiple wavelength detection with modern chemometric methods for data analysis, a third alternative procedure was described . The procedure exploits the best characteristics of both of the other methods; use of a single chiral detector; bulk in situ derivatizations; no separation step(s). Results obtained for the determinations of EP's using visible CD detection for the four ephedrine stereoisomers complexed to Cu(II) ion, were an improvement over what was capable at that time by either the chiral chromatographic or two-detector methods.

SUMMARY OF THE INVENTION

The basis for the invention was to add specificity to the biuret test through a series of modifications: (1) replace the UV absorbance detector with a CD detector; (2) substitute the achiral tartrate ion with achiral ligand to enhance the analytical sensitivity as well as the spectroscopic selectivity towards enantimers; and (3) use multiple wavelength data detection combined with chemometric data reduction algorithms to fully characterize the peptide form. Specifically, tartrate in the bioreagent is replaced by D-histidine. An aqueous solution of the Cu(II)-D-histidine complex is effectively devitalizing agent required to qualitative identify an enantiomer and quantitatively determine its enantiomeric purities.

The initial function of the host ligand (D-histidine) is to keep the Cu(II) ion in solution in high ph values. The base line CD spectrum associated with each Cu-(chiral ligand) host complex is uniquely different. On adding peptide or protein, exchange occurs between the host ligand (D-histidine) in the analyte ligand (protein). Exchanges produce changes in the CD spectrum that are significant enough that they have the potential of becoming a reliable spectroscopic fingerprint for every individual analyte, FIG. 1. If it should happen that the analyte is available in rather large quantity, characterization based upon its direct complexation with Cu(II) is preferred over ligand exchange.

CD data are measured at multiple wavelengths rather than at a single wavelength as is done in the original Biuret test. The spectrum is measured from 400-700 nm at a resolution of 0.2 nm, meaning that there are 1,500 CD signal measurements made for each analyte, thereby enhancing the analytical selectivity and accuracy of the assay.

Once CD data is obtained, two novel mathematical procedures are used for data reduction. Both are graphical procedures that achieve the reduction of the 1,500 CD data points to a single variable that is capable of characterizing each analyte structurally.

The stock reagent of the present invention is a Cu(II)-D-histidine host complex prepared from reagent grade D-histidine hydrochloride as the host ligand and reagent grade $CuS04.5H_2O$. The $[Cu^{2+}]$ =-0.020 M and [D-histidine]=0.80 M. The pH is adjusted to 13.0 with the addition of NaOh.

KI may be added as a stabilizer, preferably at a concentration of 0.30M. Stock solutions are then stable for several weeks.

Working solutions of the host complex are prepared by diluting stock solutions by a factor of 10 using 0.10 M NaOh. A typical derivatizing agent, therefore, is a solution that is 2.0 mM in Cu(II), 8.0 mM in D-histidine, 0.1 M in NaOh, and 0.03 M in KI.

In the performance of the assay, weighed amounts of analyte are added to aliquots taken from the stock solution of the host complex dissolved in an aqueous pH 13 solution (the chiral derivatizing agent). Potassium iodide may be added to the stock as a reagent stabilizer. Cu(II) ion, is introduced as CuS04. Ligand exchange occurs instantly. The mixtures are diluted to prepare working solutions part of which are transferred to fill a spectroscopic cell with a 5.0cm path length.

Exploratory investigations were controlled by making all of the measurements on solutions that contained equimolar amounts of analyte. Even greater selectivity is achieved, however, when the analyte concentration is an extra variable. That is controllable by using the same *mass* for every analyte. This is a significant move for a QC laboratory in that measuring the same mass for any and all analytes is an action very easily automated.

Complexation with Cu(II) ion in strong aqueous base, combined with visible range circular dichroism detection can be used to quantitatively differentiate among the L-enantiomers of GG, GA, GY, AG, AA, AY, YG, YA, and YY dipeptides, and the D-enantiomer of GA. Using ellipticity data at all (n=1500) wavelengths in the measured spectra, and two novel data reduction procedures, quantitative determinations are made of the compositions of binary mixtures. For mixtures made with the L-GA and D-GA enantiomers, the accuracy of the measured enantiomeric purities is better than 0,17 over the 1-48% range for the minor component. The method has considerable potential for use in quality control of peptide and protein biotechnological drug forms.

The combination of chiral ligand exchange Cu(II) complexes in aqueous base with circular dichroism spectropolarimetric detection provides excellent avenues to validate the chirality properties of oligopeptides and proteins. The method is quick and

simple and has the potential for development into an automated, routine procedure for quality control applications. Target analytes used for this first study of a protein system are human, porcine, and bovine insulins prepared by different procedures and obtained from different sources, production lots, and manufacturers. The analytical specificity of the test makes the method a potentially useful technique for validating the chirality properties for many peptide and protein forms.

The general spectroscopic method can be applied to validating aminoacid sequences in peptides and protein fragments with a view to its becoming a routine procedure with which to characterize biotechnology drug products. The tripeptides are the L-enantiomers of GGA, GGH, GGI, GGL, GGP, GHG, LGG, and YGG. The simple procedure calls for their complexation with Cu(II) ion in strong aqueous base. Binding the first three residues in the sequence, beginning at the amine terminus, completes the coordination sphere of the Cu(II) ion, so duplication of the initial sequence from peptide to peptide could be an important limiting factor in determining the extent of differentiation that is possible. The analytical focus is the selectivity associated with the chirality properties of the peptides. Detection is by circular dichroism operating in the visible range.

The procedure is done in bulk media and involves ligand exchange between the analytes and the D-histidine ligand complexed with Cu(II) ion in strong aqueous base. The metal-histidine complex functions as a chiral derivatizing reagent. Reactions with analytes produce diastereoisomers with chiral properties unique enough that they can be individually identified by their visible range CD spectra. the advantages to using D-histidine in the complex are that it serves to solubilize the Cu(II) in strong base; it produces a very stable analytical reagent with excellent selectivity; it allows for the use of much smaller amounts of analyte for an assay, which is significant when trial drugs are prepared in only very small amounts, and it functions as a common denominator against which every analyte can be compared and a single library of comparable CD spectra created.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 4 is a Visible CD spectra for the Cu(II) complexes of the nine chiral dipeptide. The most similar spectra are for GY and YY ligands especially at wavelengths shorter than 625nm.

FIG. 5 shows Correlation plots of ellipticity for the Cu(II)-(GA) complex versus the ellipticities for the analogous complexes with G(D)A, GG, AG, GA, and AA.

FIG. 6 shows Correlation plots of ellipticity for the Cu(II)-(GA) complex versus the ellipticities for the analogous complexes of YG, YA, GA, GY, YY, and AY.

FIG. 7 shows Correlation plots of ellipticity for the Cu(II)-(GA) complex versus the ellipticities for enantiomeric mixtures of percent compositions: (a) 0; (b) 97; (c) 95; (d) 90; (e) 65; (f) 52; (g) 50; and (h) 100 or pure G(D)A.

FIG. 8 shows Correlation plots of ellipticity for the Cu(II)-(GA) complex versus the ellipticities for 5 % mixtures of the GA complex plus: (A) G(D)A; (B) GG; (C) AG; and (D) AA.

FIG. 9 depict Spin Plots® for the presentation of wavelength (x-coordinate); spectral data for the GA complex, (y-coordinate); and spectral data for (A) the AG, and (B) GA complexes. The lines P1, P2, and P3, are the principal component axes from the PCA solutions.

FIG. 10 shows Linear correlation plots of PC2 values versus the % "chemical impurity" for all nine dipeptides added to GA.

FIG. 11 is a CD spectra for the series of mixtures of GA with YA. Spectra

correspond with YA impurity percentages of (A) 0; (B) 1; (C) 3; (D) 5; and (E) 10.

FIG. 12 is a Visible CD spectra for the Cu(II) complexes of: (A) GGA; (B) GGH; (C) GGI; (D) GGL; (E) GGF; (F) GHG; (G) LGG; and (H) YGG. Similarities are greatest for the GGA, GGI, GGL, and GGF complexes over the entire wavelength range.

FIG. 13 depicts Correlation plots of ellipticity for the Cu(II)GGA complex versus the ellipticities for the analogous complexes with equimolar amounts of: (A) GGA; (B) GGH; (C) GGI; (D) GGL; and (E) GGF.

FIG. 14 depicts Correlation plots of ellipticity for the Cu(II)GGA complex versus the ellipticities for the analogous complexes with equimolar amounts of (A) GGA; (B) GHG; (C) LGG; and (D) YGG.

FIG. 15 depicts Correlation plots of ellipticities for the Cu complexes of L-GHG, L-YGG, and L-LGG versus ellipticities for 5% racemic mixtures with GhG, yGG, and lGG. In each case line A is for the L-enantiomer against itself and line B is for the L-enantiomer against the mixture.

FIG. 16 depicts Correlation plots of ellipticity for the Cu(II)-(GGA) complex versus ellipticities for 5% chemical mixtures with GGH, GGI, GGL, and GGF. In each case line A is for GGA against itself and line B is for GGA against the mixture.

FIG. 17 depicts Correlation plots of ellipticity for the Cu(II)-(GGA) complex versus ellipticities for 5% chemical mixtures with GHG, LGG, and YGG. In each case line A is for GGA against itself and line B is for GGA against the mixture.

FIG. 18 depicts Spinning Plots® for the presentation of wavelength (x-coordinate); spectral data for the GGA complex (y-coordinate); and spectral data for (A) GGA; (B) GGH; (C) GHG; and (D) LGG complexes. The lines P1, P2, and P3, are

the principal component axes from the PCA solutions. Dark and light areas distinguish the front four quandrants of the cube from the rear four quadrants.

FIG. 19 shows Linear plots of the percent chemical impurity versus the eigenvector P22 for the Cu complex of GGA at a concentration of 8.0mM spiked with increasing amounts of (A) GGH; (B) GGF; (C) LGG; and (D) YGG.

FIG. 20 shows the Visible range CD spectra for: (A) the Cu(II) complex with D-histidine; and the mixed Cu(II) complexes of D-histidine with (B) bovine insulin; (C) human insulin; (D) porcine insulin; and (E) human Lyspro insulin.

FIG. 21 shows Correlations of the full spectral data for the copper-D-histidine host complex (x-axis) against the means for the corresponding data for (A) bovine; (B) human; and (C) human Lyspro insulins. The "wrap-round" nature of the plots suggests that there is 3-dimensional character to the correlations.

FIG. 22 shows Correlations of the CD spectral data for the mean Lilly human insulin mixed complex (x-axis) against the mean spectral values for (A) porcine; (B) human Lyspro; and (C) bovine insulins. Apart from human vs. human, only the human vs. porcine correlation approaches linearity for which the regression equation is ($y = 1.201 + 1.021x$, $R = 0.998$).

FIG. 23 are Spinning plots for the spectral data for (A) human Lyspro and (B) bovine insulins. The variables are the wavelength (x-axis), the full spectral data for the host complex (y-axis), and the full spectral data for the mixed complex (z-axis). P1, P2, and P3 are the principal component axes from the factor analysis.

FIG. 24 shows Zero order CD spectra for mixed complexes of D-histidine with: (A) intact bovine insulin; (B) bovine insulin A-chain; (C) bovine insulin B-chain; (D) an equimolar mixture of the A- and B-chains; and (E) the average value of the sum of the

spectra for the A-chain and the B-chain.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

In the performance of the assay according to the present invention, weighed amounts of analyte (5-15mg) are added to 1.50mL aliquots taken from a stock solution of the host complex dissolved in an aqueous pH 13 solution (the chiral derivatizing agent). Potassium iodide (30mM) is added to the stock as a reagent stabilizer. Cu(II) ion, introduced as $CuSO_4$, and host ligand concentrations in the stock are 20mM and 80mM respectively. Ligand exchange occurs instantly. Analyte concentrations must not exceed the Cu(II) concentration. Mixtures are diluted by a factor of 10 with 0.1M.

Exploratory investigations were controlled by making all of the measurements on solutions that contained equimolar amounts of analyte. Even greater selectivity is achieved, however, when the analyte concentration is an extra variable. That is controllable by using the same *mass* for every analyte. This is a significant move for a QC laboratory in that measuring the same mass for any and all analytes is an action very easily automated.

CD spectra were measured using a Jasco 500-A automatic recording spectropolarimeter coupled to an IBM compatible PC through a Jasco IF-500II serial interface and data processing software. Experimental parameters were: wavelength range 400-700nm; sensitivity 100mdeg/cm; time constant 0.25s; scan rate 200 nm/min; and temperature ambient.

Apart from the obvious advantages over MS and chromatography, which are not direct methods, other advantages that pertain to the CD method are speed (<5 minutes/test), ruggedness, simplicity in performance, accuracy, reproducibility, and the handling of safe, stable, non-toxic reagents.

Data Reduction:

Single wavelength absorbance detection is insufficient information with which to identify a compound and, depending upon the signal to noise ratio, it can also fail to provide the necessary analytical accuracy in a determination. It is virtually impossible to determine an EP greater than 90% using single wavelength absorbance data even after

chromatographic separation of the enantiomers. In the regulatory control of pharmaceutical products, values better than 99.0% are required.

Multiple wavelength detection can solve the identification problem but it does introduce the other problem of how to conveniently handle so much data. A high priority has been given in the analytical community to deriving practical mathematical algorithms that have the discriminatory power to achieve qualitative **and** quantitative analyses.

The 2-D Data Reduction Model:

In this treatment, the 1500 data points that make up the CD spectrum for the host complex are plotted (on the x-axis) against the analogous spectral data (on the y-axis) for the complex where the host ligand is partially exchanged with the analyte, FIG. 2.

If the host ligand and the analyte are identical, the plot is a straight line of slope equal to 1.0. If the host ligand is the D-enantiomer of an L-analyte, or vice versa, and they are equally pure, the plot is again a straight line but with a slope of -1.0. Mixtures of enantiomers of unequal concentrations still give straight lines but the lines have slopes that fall in the range of ±1.0. Enantiomeric ratios correlate directly with the calculated slopes of these lines. Using this data reduction model, EP's better than 99.8% were easily measured.

Analytes that are chemically different from the hostligand produce characteristic non-linear plots that can be used for identification purposes, FIG. 2, but not for purity determinations.

The 2-D model has special value in that it will detect the presence of chemical impurities in a chiral product, on the basis of a loss of linearity, and is able to measure EP's with greater accuracy than has ever been accomplished before.

The single datum (variable) that emerged from the data reduction is the number that gives the slope of the line in the 2-D spectral plot.

The 3-D Data Reduction Model:

The 3-D model is an elementary expansion of the 2-D model in which wavelength is added as the third variable. Related 3-D plots are presented pictorially as Spinning

Plots®, FIG. 3.

Data reduction of the 3-D data is done by Factor Analyses, or Principal Component Analyses (PCA) of the Spinning Plot® data. The result is a matrix of 12 eigenfactors, Table 1. Principal Components that comprise the model are fitted by three eigenvalues and nine eigenvectors (PC coordinates). The resultant principal components P1, P2, and P3 have been added to FIG. 3. Data reduction has effectively taken 1500 CD data points and reduced them to just 12 variables.

Analytical Identifications:

From several studies made on a number of peptide and protein systems, the PCA evidence is clear that most of the variabilities among different analytes are limited to no more than 4 of the 9 eigenvectors, i.e. PC22, PC23, PC32, and PC33, Table 2.

Table 1: Principal Component Analysis of the CD Spectral Data for the Complex Formed When D-Histidine is Exchanged with 0.96mM L-alanyglycine.

	P1	**P2**	**P3**
Eigenvalues	2.2647	0.7346	0.0007
Eigenvectors	PC11=-0.54608	PC21=0.83495	PC32=0.06825
	PC12=0.58888	**PC22=0.44054**	**PC32=-0.67760**
	PC13=0.59583	**PC23=0.32984**	**PC33=0.73225**

* Values in bold type show the greatest sensitivity with respect to the identity of the analyte

TABLE 2 : Variation of Most Sensitive PC Values. Mass of Analyte = 10.0mg. (Standard Deviation on PC values is ± 0.002)

ANALYTE	PC22	PC23	PC32	PC33
DIPEPTIDES				
alanylalanine AA	0.61813	0.13116	- 0.53700	0.79051
alanylglycine AG	0.64224	0.09811	- 0.51030	0.79477
alanyltyrosine AY	0.69846	0.01430	- 0.436060	0.80127
glycylalanine GA	0.46494	0.30777	- 0.66336	0.74206
glycyl(D)alanine G(D)A	0.16171	0.57548	0.77290	- 0.57895
glycylglycine GG	0.42950	0.34106	- 0.68392	0.72752
glycyltyrosine GY	0.56357	0.19935	- 0.58973	0.77664
tyrsoylalanine YA	0.25849	0.49507	0.75322	- 0.64094
tyrosylglycine YG	0.29769	0.46040	0.74112	- 0.66365
tyrosyltyrosine YY	0.51030	0.25913	- 0.63241	0.75925
TRIPEPTIDES				
GGA	- 0.18286	0.90459	0.66270	0.40274
GGH	- 0.53847	0.83609	0.60213	0.29475
GGI	- 0.63506	0.77202	0.53978	0.41922
GGL	- 0.68830	0.72530	0.48586	0.44702
GGP	- 0.11093	0.88764	0.67635	0.39696
LGG	- 0.02795	0.95025	0.71043	- 0.19945
YGG	0.17842	0.89443	0.66022	- 0.41783
GHG	0.26282	0.92041	0.65294	- 0.39052
GGG	0.58033	0.18668	- 0.57453	0.78371
NEUROPEPTIDES				
DADLE	- 0.01355	0.71678	0.78638	- 0.42301
DAGO	- 0.00098	0.70862	0.74621	- 0.46920
DSLET	0.02654	0.68858	0.79064	- 0.45806
DTLET	- 0.02556	0.72591	0.78316	- 0.41276
DynorphinA (1-9)	0.31022	0.47799	0.74661	- 0.65760
DynorphinA (1-11)	- 0.07004	0.75670	0.77995	- 0.36469
Dynorphin A(1-13)	0.00659	0.70258	0.79073	- 0.43926
DynorphinA (1-13)NH2	- 0.24208	0.85195	0.74292	- 0.14503
DynorphinB (1-13)	0.31640	.45077	0.73678	- 0.67100
Met-enkephalin	0.78708	- 0.17370	- 0.24973	0.78665
B-endorphin	- 0.06949	0.75799	0.77775	- 0.36596
INSULINS				
Human Insulin	0.78851	- 0.60425	0.31377	0.55442
Human Lyspro	0.73437	- 0.67821	0.42275	0.48814
Porcine Insulin	0.79510	- 0.59239	0.29464	0.56464
Bovine Insulin	0.82594	- 0.50719	0.16287	0.63266
Bovine Chain A	0.01021	- 0.69856	0.77732	0.45565
Bovine Chain B	- 0.14041	0.91557	0.67588	0.36677

Analytical Determinations:

The PCA analog of the conventional linear Beers' Law correlation of absorbanc vs. Analyte concentration is the linearity in the plot of PC23 vs. Analyte concentration, FIG. 4. With this information, it is possible not only to detect a chemical impurity but also to determine the proportionate amount.

The outcome of the data reduction strategies is that whereas the 3-D data reduction model fails to detect an enantiomeric impurity, but does allow for the determination of chemical impurities, the 2-D model measures EP's with extraordinary accuracy, but detects chemical impurities at a qualitative level only. The two models are complementary in determining the chemical and the enantiomeric impurity levels of peptides and protein drug forms in a simple spectroscopic QC analysis.

The three levels of spectral discrimination represented by FIGS. 1-3 are obtained from the same spectroscopic data set.

CHIRAL PROPERTIES OF PEPTIDES.

The object of this the first investigation of a simple peptide system is to prepare a series of chiral Cu(II)-dipeptide metal complexes and to determine the degree of analytical selectivity that is possible using visible range CD spectrophotometric detection. Prospects for obtaining a quantitative measure of both enantiomeric and chemical impurity levels are discussed at some length. The order of residues in short peptides is crucial to the extension of the study to peptides and proteins as a whole because coordination of the latter to Cu(II) involves the first three aminoacids that make up the amine-terminus. If there is no CD selectivity associated with changes in the initial sequence, the method has no value in the study of oligopeptides and proteins where too frequently the same initial sequence is common to several potential analytes.

Dipeptides chosen for the study have sequences that are permutations of only three L- and one D-aminoacid monomers. The molecules have no ternary structure, so the sequence variation is the only factor that will affect the selectivity. The experimental procedure is a simplification of the method that was used to measure EP's for ephedrine

mixtures. Data reduction and spectral differentiations are done using variations on standardized mathematical algorithms.

Experimental:

Chemicals: Nine of the dipeptides used in the study are analogs of glycine (G), L-alanine (A), and L-tyrosine (Y). Each aminoacid occupies a position at either end of the "peptide chain", viz. GG, GA, and GY; AG, AA, and AY; and YG, YA, and YY (the amine terminus is listed first). The last peptide is the D-alanine enantiomer of glycylalanine, abbreviated to G(D)A. All ten were supplied by Sigma Chemical Co. which reported an EP in excess of 99.8%. Reagent grade D-histidine, used to calibrate the CD scale, was also Sigma Chemical Co. product. Reagent grade Cu(SO4)2. 5H2O was obtained from Fisher Scientific.

Solution Preparations: Stock solutions were prepared for each of the Cu(II)-dipeptide complexes in pH 13 aqueous solution in which the Cu2+ concentration was always 0.020M. The range of di-peptide concentrations was 0.005; 0.01; 0.02; 0.04; and 0.08M respectively. KI at a concentration of 0.03M was added as a stabilizer. Working solutions were prepared by diluting the stocks by a factor of ten with 0.10M NaOH.

For mixture analyses, GA was arbitrarily selected as the enantiomerically pure "reference" substance. Predetermined volume aliquots of the remaining nine dipeptides were added that cov-ered the range of "impurity" from 1-48%, prior to dilution with NaOH.

The chemistry of the derivatization reaction is a simple chiral variation of the classical biuret "color reaction" for the determination of total serum proteins in which the reagent is a solution of [Cu(II)] = 2.0mM and [tartrate] = 8.0mM in 0.10M NaoH. Racemic tartrate is the solubilizing ligand for Cu(II) and is completely exchanged by protein in the test. The chemistry for the re-action is well understood and relatively uncomplicated. Determinations are based upon absorbance spectrophotometric detection. Being relatively insensitive and not sufficiently selective, the biuret reaction is no longer the method of choice for scrum proteins.

Measurements:

CD spectra were measured using a Jasco 500-A automatic recording spectropolarimeter coupled to an IBM-compatible PC through a Jasco IF-500 II serial interface and data processing software. Experimental parameters were: wavelength range 400-700nm; sensitivity 100 mdeg/cm; time constant 0.25s; scan rate 200nm/min; pathlength 5.0cm; temperature ambient. Calibration of the day to day reproduciblity of the system was done by measuring the CD spectrum for a reference solution of Cu(II)-D-histidine in which the [Cu(II)] = 2.0mM and the [D-histidine] = 8.0mM. Statistical data for reproducibilities of the maximum ellipticity values measured at wave-lengths 487nm and 682nm were 7.42±0.07 mdeg and -214±0.60 mdeg respectively.

Results and Discussion:

Cu(II)-amide and histidine complexes: The local microsymmetry of the Cu(II) ion in aqueous solution is essentially square-planar due to Jahn-Teller distortion, or axial elongation, of the local octahedral symmetry generally adopted by most first row transition metal ions. Complexation serves to keep the Cu(II) ion in solution at high pH conditions.

At pH=13, D-histidine and the amide-nitrogen protons are fully ionized, which simplifies the competitive nature of the complex formation equilibria. D-histidine, the ligand used for instrument calibration, binds via the amine N-atom, the carboxylate functional group, and a pyrimidine N-atom in an equatorial three-coordinate arrangement. Stoichiometry for the complex is 1:1.

Complexing a peptide to Cu(II) at high pH involves first attachment through the N-atom of the terminal amine followed by ring closure(s) through bonding with the N-atoms of successive amide bonds until maximum thermodynamic stability is achieved. Side chain substituents on the aminoacid residues lie out of the coordinate plane and are factors only in inter- and intramolecular interactions within the inner coordination sphere, unless a potential Lewis base is present, e.g. a histidine residue. Axial positions might be occupied by hydroxide ions which is the only potential complicating feature of stoichiometry of the generic metal-peptide, (MP)n , equilibrium.

By analogy with the Cu(II)-D-histidine equilibrium reaction, the stoichiometry

of the Cu(II)-dipeptide complexes are also believed to be 1:1. If the only purpose of the study were to develop analytical selectivity, the question of the stoichiometry of the metal-dipeptide complexes is moot. If the stoichiometry were to change from one ligand to another, the selectivity might very well be enhanced. Conventionally it is only when making an analytical determination that knowledge of the stoichiometry a prerequisite.

CD activity in the visible range for chiral Cu(II) complexes is a result of disymmetric perturbations of the ground and excited state ligand field orbitals by the chiral ligands. Bands in the UV range, attributable to only the chirality in the ligands, bound and unbound, are typically very in-tense but quite insensitive to the environment of the coordinating metal ion. The lack of selectivity is the major reason for not exploiting the obvious analytical sensitivity that is inherent in the intense UV bands.

Visible CD spectra for Cu-complexes: Spectra for copper complexes with D-histidine (8.0mM) and all ten dipeptides (8.0mM) are shown in FIG. 4. The spectrum for the D-histidine complex is biphasic and dominated by the intense negative band that maximizes at 682nm. For the achiral Cu-(GG) complex, the spectrum is coincident with the baseline. Spectral variations are sufficient to differentiate among the remaining dipeptides with the possible exception of GY vs YY which differ only at wavelengths longer than 625nm.

Indicative of the sensitivity of the detector to the monomer sequence are the differences between spectra for the inverse pairs, GA/AG, GY/YG, and AY/YA. The residue whose N-atom is directly involved in the ring-closure step has the greater effect on the magnitude of the spectral changes. Although G- is achiral, its presence and relative position affects the CD spectra quite dramatically which is good reason to believe that specific interligand interactions occur within the first coordination sphere of the complex.

Factors that Affect the Selectivity:

Solution factors that contribute to the observed spectral differences are: total analytical concentrations; (in)constancy of the stoichiometries and relative stabilities of the complexes; the rotatory strength for each complex. The first factor is normalized by staying with the same analytical concentration for every ligand. Based upon spectral

comparisons made at all five ligand concentrations, it was concluded that the best discriminations are observed at a [ligand] = 8.0mM.

At ligand concentrations less than [Cu(II)] = 2.0mM, the metal ion is precipitated almost immediately. Stable stock solutions require the ligand to be in excess, cf. tartrate to Cu(II) ion in the biuret test is 4:1. Using CD data for stock solutions in which the [ligand] 2.0mM, and the graphical procedure developed by Newton and Arcand for absorbance data for the Ce(SO4)n complexation reaction. For the metal-peptides, maximum ellipticity data were plotted against [peptide]-n for different values of n, at constant [Cu(II)]. The stoichiometry is given by the only value of n that produces a straight line plot, in this instance n = 1, in keeping with the assumption made earlier that peptide and D-histidine complexations are analogous. The mean value for the 1:1 formation constants for all nine CD-active complexes, calculated by an iterative procedure , is 2.30x109 ±0.4x109 which compares very well with literature values that are on the order of 1010at a pH > 12 for related anions.

Because of the strong similarities in stabilities and CD band intensities, the conclusion is reached that rotatory strengths are also of the same order of magnitude for all complexes. Spectral selectivity is probably dominated therefore by the very specific chiral-chiral interactions that occur between neighboring coordinated aminoacid ligands.

Alternative Algorithms for Data Reduction and Enhancing Selectivity: Most algorithms deal with data measured at the wavelength of the maximum signal only; unless a chemometrics approach is employed. The intent of the algorithms described here was to take ellipticity data at all 1500 wavelengths and, by novel mathematical procedures, reduce the data to just one variable (or factor) upon which selectivity decisions are made. Having a simple numerical means for making selectivity judgments is superior to relying upon subjective graphical superpositions of the CD spectra. If that same numerical factor were to correlate linearly with ligand concentration, then quantitative differentiations might also be made. Resultant analytical determinations will be more accurate since experimental uncertainties are significantly reduced when 1500 data points are used rather than only one.

2-D Data Reduction Algorithm:

For this elementary data reduction procedure, GA is arbitrarily assigned the status of an enantiomerically pure standard reference material. In a pharmaceutical context, GA might represent a commercial drug product. The others fill the roles of potential "chemical" and "enantiomeric" impurities.

The simple concept behind the data reduction is to plot the 1500 data points for the 8.0mM GA spectrum (on the x-axis) against analogous data for 8.0mM solutions for each of the others (on the y-axis). To get a baseline reference check for the absolute enantiomeric purity of GA, its CD spectrum is plotted on both axes. The resultant is a straight line of unit slope and zero intercept. Spectra for the remaining nine dipeptides are plotted against GA in FIGS. 5 and 6.

Only the plot for the enantiomeric G(D)A complex shows a similar linear behavior with a slope close to -1.0, FIG.5. The extreme similarity implies a chemical correspondence between the two. For GG the slope of the line is zero. All other plots are decidedly non-linear and very distinct from one another. Qualitative differentiation among enantiomerically pure forms of the dipeptide is elementary at least when the number of possibilities is limited to a small closed set, as they are here. A small remnant of the ambiguity that was seen in differentiating GY from YY by their zero order CD spectra still remains in these coordinate plots.

(a) ENANTIOMERIC MIXTURES OF GA+G(D)A. If both enantiomers are of equivalent purity, slopes will be ±1.0 respectively. The slope of the line in FIG. 5 for the parent G(D)A enantiomer is - 0.998, which transforms to roughly a 0.05 % impurity of the L-enantiomer, if the assumptions made that the EP for the L-enantiomer is absolute, and the only impurity is the L-enantiomer are correct. As G(D)A is added to GA in increasing amounts, the slopes of the correlation lines de-crease until a value of zero is reached for the racemate, FIG. 7. Judging by regression coefficients, there is no significant loss of linearity compared to the reference baseline. Retention of linearity is categorical proof that the "impurity" is the enantiomer of the same chemical material.

Enantiomeric excesses, defined as:

{[Cu(II)(GA)] - [Cu(II)(GA)1-x(G(D)A)x]} / {[Cu(II)(GA)] +

[Cu(II)(GA)1-x(G(D)A)x]}

and calculated from the correlation slope for each mixture, are in excellent agreement with the experimental values for prepared mixtures, even at the extremes of 52% and 99%, Table 3. Imprecisions in the calculated EP's, based upon 3-5 repeat measurements, are improved by almost a factor of 10 over the results obtained from the analyses of binary ephedrine mixtures in which a chemometric analysis method was applied to data at 5 wavelengths.

Table 3: Determination of Enantiomeric Purities for Prepared Binary Mixtures of GA and G(D)A.

% GA in Prepared Solution	Regression Slope	Regression Coefficient	% GA Calculated
100	1.0	1.0	100
99	0.9837	0.99994	99.05 ± 0.15
97	0.9438	0.99995	97.01±0.19
95	0.8998	0.99994	95.03±0.03
90	0.8016	0.999898	90.05±0.04
65	0.2975	0.9995	64.96±0.08
52	0.0400	0.96388	52.13±0.13
50	0.0040	0.3391	50.34±0.17
0	-0.9989	1.0	0.05±0.05

Table 4: Principal Component Values Calculated for the GA versus AG System:

		PC1	PC2	PC3
Eigenvalues		+1.9681	+1.0008	+0.0311
Eigenvectors	nm	-0.70725	-0.00007	+0.70697
	GA	+0.52044	-0.67686	+0.52058
	AG	+0.47848	+0.73611	+0.47874

In summary, therefore, the 2-D algorithm has effectively reduced the 1500 spectral data points to one number (the correlation slope), from which EP's can be determined with excellent accuracy. Identifications of potential chemical impurities are possible but their analytical determinations are not easily done.

(b) GA +"CHEMICAL IMPURITY" LEVELS FOR ALL OTHER DIPEPTIDES.

The question with respect to "chemical impurities" that needs to be addressed is not how great the absolute differences are between the curve for an 8.0mM solution of GA and curves for the other ligands at equal concentrations, Figs. 5 and 6, but rather, are the differences sufficient enough to identify and quantitate "chemical impurities" when these amount to only a few percent of the composition of a binary mixture ? This can be answered in the following way.

In the hypothetical large scale manufacture of the "drug" GA, "chemical impurities" might be G- and A- monomers, and GG, AG, or AA dimers. Monomers do not chelate to metal ions, and are not thermodynamically competitive with dimers in binding, so they do not contribute to the visible CD spectrum. If monomers are present, the actual amount of GA in the weighed sample aliquout is lowered slightly, and the resultant slope is less than 1.0. The correlation line will still be straight, so a monomer might be mistaken for either achiral GG or the enantiomer G(D)A.

Low amounts of dimeric impurities, on the other hand, are easily detected by the obvious split-ting of the correlation lines, compared with the baseline for GA vs.GA. Splitting is accompanied by a change in the virtual slope of the line, FIG. 8. The extent of each split has been amplified on the y-axis in FIG. 8 by omitting data at the shortest wavelengths. Modeling the "characteristic" curves of Figs. 5,6, and 8 to determine the amount of a given dimeric impurity is not something that can be done using linear mathematical models on these data.

3-D Data Reduction Algorithm:

In the 2-D presentations of Figs. 5 and 6, wavelength is an im-plied variable. For the evolution of the 3-D algorithm, wavelength is the third dimension. Since the experimental parameter measured in CD detection is an absorbance difference, observed

signals are positive, negative, or zero. When two CD spectra are plotted against each other, four sign combinations are possible at any wavelength. Repeats of coordinate points, e.g. zero crossover points, can occur at wavelength values that are not adjacent to one another in the spectra. When that occurs 2-D plots "wrap around" and become three dimensional. In retrospect what are observed as 2-D plots are projections of the 3-D plots on to the x-y coordinate plane, which explains why some of the plots in Figs. 6 and 7 appear to have a 3-D character. The added value of the third dimension is that there should be an increase in the overall analytical selectivity.

The algorithm used for the visual presentation of the three parameter plot was Spinning Plot® which is an integral part of a number of commercially available statistical analyses software packages. Representative 3-D plots are shown for wavelength (nm) vs. GA vs. AG (FIG. 9A) and GA (FIG. 9B). The latter is included to show the 3-D nature of the "baseline" plot of GA against itself. In qualitative terms visual differentiations are easier than they are for the 2-D plots. Most important, perhaps, is the certain, and final, distinction between GY and YY.

Factor Analyses of Spinning Plot® Data:

For the derivation of a quantitative mathematical algorithm, data reduction was done using a Principal Component Analysis (PCA) procedure . Eigenvalues and eigenvectors for the three principal components, PC1, PC2, and PC3, calculated for the GA/AG combination plot are given in Table 4. Quantitative spatial projections of these same principal components are superimposed on the coordinate axes of FIG.9.

Of the twelve eigenfactors, the one with the most sensitivity to changes in the identity of the analyte is the PC2 value in row 3 of the eigenvector matrix, shown in bold type in Table 4. For GA plotted against itself, PC2 is 0.37286. Corresponding values for GA plotted against spectral data for the remaining ligands in 8.0mM solutions are: AA (0.10644); AG (0.73611); GG (0.99917); YA (0.69055); GY (0.09830); YG (0.53344); AY (- 0.18498); YY (0.17204); G(D)A (-0.36786). There is a very clear distinction between the correlations for the achiral GG and the enantiomeric G(D)A peptides. Standard deviations calculated for spectral data from 3-5 independent repeat

measurements are ±0.002, substantiating the earlier claim that differentiations can be arrived at by a simple mathematical inspection of the 3-D plots. The small difference between the absolute PC2 values for GA and G(D)A is consistent with the small difference in the slopes of the correlation lines in the 2D treatment, implying that their EP's are not exactly equivalent.

In summary, therefore, the 3-D algorithm has effectively reduced the 1500 spectral data points to one number, PC2, with which potential chemical impurities can be qualitatively identified.

Quantitative determination of the amount of a "chemical impurity", that is on the order of 1-20% of the amount of GA, followed quite simply from the PC2 determinations. PC2 values calculated for each binary mixture, including the enantiomeric pair, correlate linearly with the percent impurity values, FIG. 10. With the exception of GG and G(D)A, correlation slopes give analytical sensitivities that are at least ten times more accurate than analogous plots of ellipticity values measured at a maximum wavelength plotted against concentration. Why this is so is easily understood from FIG. 11, in which CD spectra are plotted as a function of the percent YA impurity. With a best possible resolution of ±2.0mdeg for the CD instrumentation used in this study, there are cases among these analytes where the S/N is of insufficient quality to obtain a determination at all.

The slope of the PC2 vs percent impurity line of YA in FIG. 10, which approximates to the mean value for all of the slopes, is more than two times the ±0.002 SD in the means for PC2 values, which means that impurity levels as little as 1-3% can be measured with confidence.

Where the 2-D method succeeded in providing accurate values for EP's, the 3-D method provides for a quantitative measure of the non-enantiomeric chiral impurities.

Summary and Application of the Method:

By the simple chiral modification of the biuret reagent, combined with two novel data reduction algorithms for the handling of visible CD data, a potentially useful QC regulatory procedure for peptides, oligopeptides, and proteins has been developed. It fits QC circumstances, where the objective is to test the purity of a single chiral substance

where the amounts of the potential impurities are small, very well.

A typical procedure begins with the measurement of the visible CD spectrum for the Cu(II) complex of the chromatographically purest available form of the substance being regulated. Data are archived in a PC file on-board the spectrometer and are updated each time the reference material is measured. Spectra for aliquots taken from each newly manufactured product lot are plotted against the standard and successive on-screen visual comparisons are made.

Deviation from a slope of 1.0 in the 2-D test is an instant indication that the purity is less than that of the reference standard. If the 2-D linear regression coefficient indicates an enantiomeric "impurity", the EP is calculated from the regression slope. Separation in the correlation line for the standard reference material gives instant recognition that a "chemical impurity" is present whose identity is confirmed by the PC2 value calculated from the Spinning Plot® algorithm. The percent impurity is calculated from the correlation slope of the PC2 vs impurity line.

The method is quick, rugged, uses stable inexpensive reagents, requires no specific precautions, and a minimum of technical expertise for a potential operator. Derivatization reactions are instantaneous and data collection is done in a matter of minutes. Spectral data are stored on an on-board computer that is programmed to perform all the mathematical comparisons and quantitative analyses in situ.

All of these advantages point to a very satisfactory and very competitive routine alternative to chromatographic and mass spectrometry methods for the quality control of small peptides.

TRIPEPTIDE DISCRIMINATIONS USING CIRCULAR DICHROISM DETECTION.

A major new frontier in the pharmaceutical industry is the focus on the therapeutic properties of peptide and protein drug forms. Because the number of chiral centers has virtually no limit, the magnitude of the chirality regulatory control problem is increased almost exponentially. Since derivatizations will not produce a single

diastereoisomer, even the very best chiral chromatographic methods face what are probably insurmountable challenges unless the peptides are first cleaved enzymatically .

Problems that are associated with chirality detection also increase. The total CD signal for a metal-peptide complex is not determined by just the number and sequence of chiral centers in the primary peptide structure. It also includes contributions from longer range chiral interactions between side-chain substituents that modify the ternary structure when peptides are coordinated to metal ions. Experimental conditions must be very carefully controlled otherwise these very pH-sensitive structural modifications would give false information about the analyte to the detector. On the other hand the simple accumulation of these additive chiral properties could conceivably pro-duce a level of analytical selectivity that is unmatched by other detectors and might even approach specificity. Enzymatic cleavage followed by CD detection is also an option. What chirality detection contributes that the others do not is a direct look at the enantiomeric form. This ability will in-crease in value as long as manufacturers continue to use D- for L- enantiomeric substitutions as a strategy in peptide drug design.

Eight tripeptides were chosen for the study. Common to all eight are two glycine residues which occupy positions 1,2-, 1,3-. and 2,3- in the sequence. The remaining residues are L-enantiomers of aliphatic and aromatic aminoacids. The tripeptides have no stable ternary structure to speak of, so variability in the sequence is really the only parameter affecting the chiral response of the CD detector. The order of residues in short peptides is crucial to the extension of the study to peptides and proteins as a whole because coordination of the latter to Cu(II) involves the first three aminoacids from the amine-terminus. If there is no CD selectivity associated with changes in the initial sequence, the method has no value in the study of oligopeptides and proteins where too frequently the same initial sequence is common to several potential analytes.

The experimental procedure is a combination of the methods that were used to discriminate among related dipeptides and insulins and to measure EP's for glycyl-L-alanine and ephedrine mixtures. Data reduction and spectral differentiations are done using variations on standardized mathematical algorithms and principal component analysis (PCA).

Experimental:

Chemicals: Tripeptides used in the study were glycylglycyl-L-alanine (GGA), glycylglycyl-L-his-tidine (GGH), glycylglycyl-L-isoleucine (GGI), glycylglycyl-L-leucine (GGL), glycylglycyl-L-phenylalanine (GGF), glycyl-L-histidylglycine (GHG) and its D-enantiomer (GhG), L-leucyl-gly-cylglycine (LGG) and its D-enantiomer (lGG), and L-tyrosylglycylglycine (YGG) and its D-enantiomer (yGG). All eight L-enantiomers were supplied by Sigma Chemical Co. which reported an EP in excess of 99.8%. The D-enantiomers GhG, lGG, and yGG of GHG, LGG, and YGG were prepared by Multiple Peptide Systems (MPS), San Diego. Certificates of Analysis described them as unpurified off white powders. Percent purities as determined by RP-HPLC analyses were re-ported as 86.57; 99.10; and 97.86 respectively. The low value for GhG is explained as being due to two elution peaks that correspond to the same compound. The percentage is based on the relative area of the first peak which corresponds with most of the material eluting with the void volume peak. The second peak is related to the hydrophobicity of the molecule that causes it to stick to the column to be eluted later. D-histidine was also a Sigma Chemical Co. product with an EP reported at better than 99.8%. Reagent grade CuSO4.5H2O was obtained from Fisher Scientific.

Solution Preparations:

The chemistry of the derivatization reaction is a simple chiral variation of the classical biuret "color reaction" for the determination of total serum proteins in which the reagent is a solution of [Cu(II)] = 2.0mM and [tartrate] = 8.0mM in 0.1M NaOH. Racemic NaK-tartrate is the solubilizing ligand for Cu(II) and is completely exchanged by protein in the test. The chemistry for the reaction is well understood and relatively uncomplicated. Determinations were done based upon absorbance spectrophotometric detection. Being relatively insensitive and not sufficiently selective, the biuret reaction is no longer the method of choice for serum proteins.

In this instance, aqueous stock solutions at pH 13 were prepared for Cu(II)-D-histidine and each of the Cu(II)-L-tripeptide complexes in which the Cu2+ concentration was always 0.020M. Ligands were present at 0.080M concentrations, a four to one excess over the Cu(II) ion. KI at a concentration of 0.03M was added as a

stabilizer. Spectra were measured for working solutions prepared by diluting stocks by a factor of ten with 0.10M NaOH. Spectra for the working solutions are the bases for testing the extent of the qualitative analytical selectivity accessible to CD detection.

Quantitation tests were done on two kinds of mixtures. For the first kind, GGA was arbitrarily selected as an enantiomerically pure "reference" material. Aliquots from the stock were spiked with "chemical impurities" i.e. smaller volume aliquots of the other L-tripeptide stocks to cover the im-purity range from 1 to 10%, prior to dilution with NaOH. For the second, "enantiomeric purity" tests, aliquots of YGG, LGG, and GHG stocks were spiked with smaller volume aliquots of the corresponding D-enantiomer stocks, yGG, lGG, and GhG over the same 1-10% impurity range.

Measurements:

CD spectra were measured using a Jasco 500-A automatic recording spectropolarimeter coupled to an IBM-compatible PC through a Jasco IF-500 II serial interface and data processing software. Experimental parameters were: wavelength range 400-700nm; sensitivity 100 mdeg/cm; time constant 0.25s; scan rate 200nm/min; pathlength 5.0cm; temperature ambient.

Calibration of the day to day reproduciblity of the system was done by measuring the CD spectrum for the Cu(II)-D-histidine complex. Statistical data for reproducibilities of the maximum ellipticities measured at wavelengths 487nm and 682nm were 7.42 ±0.07 mdeg and -214 ±0.60 mdeg respectively.

Results and Discussion:

Cu(II)-peptide and D-histidine complexes: The local microsymmetry of the Cu(II) ion in aqueous solution is essentially square-planar due to axial elongation of the typical octahedral symmetry, assumed by most first row transition metal ions, by Jahn-Teller distortion. Complexation serves to keep the Cu(II) ion in solution at high pH conditions. At pH 13, D-histidine and the amide-nitrogen protons are fully ionized, which essentially eliminates competitive complex formation equilibria when partially protonated anions are present in solution at lower pH.

D-histidine, the ligand used for instrument calibration, binds via the amine N-atom, the car-boxylate functional group, and a pyrimidine N-atom in an equatorial three-coordinate arrangement. Stoichiometry for the complex is 1:1. Complexing a peptide to Cu(II) at pH > 12 involves first attachment through the N-atom of the terminal amine followed by ring closure(s) through bonding with the N-atoms of successive amide bonds until maximum thermodynamic stability is achieved. Side chain substituents on the aminoacid residues lie out of the coordinate plane and are factors only in inter- and intramolecular interactions within the inner coordination sphere, un-less a potential Lewis base is present, e.g. a histidine residue. Axial positions might be occupied by hydroxide ions which is the only feature that might complicate the stoichiometry of the generic metal-peptide, (MP)n , equilibrium.

By analogy with the Cu(II)-D-histidine equilibrium reaction, the stoichiometry of the Cu(II)-tripeptide complexes is also believed to be 1:1. If the only purpose of the study were to develop analytical selectivity, the question of the stoichiometry of the metal-tripeptide complexes is not relevant. If the stoichiometry were to change from one ligand to another, the analytical selectivity might very well be enhanced. It is only when making an analytical determination by conventional mathematical procedures that knowledge of the stoichiometry is a prerequisite.

CD activity in the visible range for chiral Cu(II) complexes is a result of disymmetric perturbations of ground and excited state ligand field orbitals by the chiral ligands. Bands in the UV range, attributable to only the chirality in the ligands, bound and unbound, are typically very intense but quite insensitive to the environment of the coordinating metal ion. The lack of selectivity is the major reason for not exploiting the obvious analytical sensitivity that is inherent in the intense UV bands.

Visible CD spectra for Cu(II)-tripeptide complexes Spectra for all eight copper-L-tripeptide complexes, in which [Cu(II)] = 2.0mM and ligand concentrations are 8.0mM, are shown in FIG. 12. Only GGH, GHG, LGG, and YGG are uniquely differentiable by their zero order CD spectra. Spectra for the histidyl-containing ligand complexes, GGH and GHG, are blue shifted compared with the Cu(II)-D-histidine complex itself which has an intense negative band with a maximum at 689nm and a

weaker positive maximum at 570nm. The magnitude of the shift is greatly dependent upon the position occupied by the histidyl residue. The sensitivity of the CD spectral response to the histidine position is a significant first result in the context of possibly sequencing short peptides by this spectroscopic method.

Of the five GGX peptides, only the spectrum for GGH is unique, which might be attributable to a special involvement of the pyrimidine N-atom in binding to Cu(II). The remaining four have but one broad negative band that maximizes around 550nm. Aromaticity in the side chain may (YGG) or may not (GGF) induce a spectral change, which with further developments, might be exploited for short range sequencing. There is ambiguity in differentiating among GGA, GGI, GGL, and GGF unless the solution concentrations are carefully controlled.

The L-leucine structural isomers can ostensibly be differentiated in a quality control context. The lack of band intensity for the LGG is a potential problem in quantitation. Although the glycyl- residue is achiral, the relative positions that it occupies affect the CD spectra quite dramatically which is good reason to believe that specific interligand interactions occur within the first coordination sphere of the complex.

It is quite clear at this point that total differentiation among all eight analytes is not possible.

Alternative Algorithms for Data Reduction and Enhancing Selectivity: Conventional algorithms typically deal with data measured at just the wavelength of the maximum signal; unless a chemometrics approach is employed. The intent of the algorithms described here was to start with ellipticity data measured at all 1500 wavelengths and, using novel mathematical procedures, reduce the data to a single variable (or factor) upon which selectivity decisions are made. Having a simple numerical means for making selectivity judgments is superior to relying upon subjective graphical superpositions of the CD spectra. Furthermore, if that same numerical factor were to correlate linearly with ligand concentration, then quantitative differentiations might also be accomplished. As a final consequence, resultant analytical determinations will be more accurate since experimental uncertainties are significantly reduced when 1500 data points are used rather than only one.

A 2-D Data Reduction Algorithm for Enhancing Selectivity:

To illustrate this data reduction procedure, GGA is arbitrarily assigned the status of an enantiomerically pure standard reference material. In a pharmaceutical context, GGA might represent a commercial drug product. The others fill the roles of potential "chemical" and "enantiomeric" impurities.

The simple concept is to plot the 1500 data points for the 8.0mM GGA spectrum (on the x-axis) against analogous data for 8.0mM solutions for each of the others (on the y-axis). To get a baseline reference check for the absolute enantiomeric purity of GGA, its CD spectrum is plotted on both axes. The correlation is a straight line of unit slope and zero intercept. Spectra for the remaining seven tripeptide complexes are plotted against GGA in FIG. 13 for the GGX sub-series and in FIG. 14 for the other sequences.

Plots are decidedly non-linear and individually distinct from one another. The ellipsoidal shapes for GGI and GGL might appear similar but the best-fit lines do have different slopes. On enlargement, however, the ellipse for GGI is seen to "fold over" on itself in a partial FIG. 8 implying a latent 3-dimensional property in these plots. The same phenomenon can be seen more clearly for the plot of the GGH analog vs. GGA in FIG.14. Differentiation among enantiomerically pure forms of the tripeptides has apparently been achieved at least when the number of possibilities is limited to a small closed set, as they are here. It should be emphasized that in order to reproduce these curves exactly the concentrations must be carefully controlled.

The only other possible correlation line of unit slope (but opposite in sign) and zero intercept is the plot of GGA vs. the D-enantiomer, GGa, if their purities are equivalent. This is a consequence of their being chemically identical. The feature that is common to all cases where spectra for chemically dissimilar compounds are correlated is splitting of the correlation line relative to the ideal reference line. Some splittings are extreme, Figs 13, 14. Conversely, if the correlation plot of the CD spectrum for a newly manufactured lot of GGA vs. the reference is linear with a slope less than one, and shows no evidence of splitting, this is evidence for the presence of the enantiomer. Splitting is instant evidence for the presence of a chemical impurity.

Non-linear plots, typical of Figs.13 and 14 do not yield easily to quantitation of

the "chemical impurities".

(a) QUANTITATION OF ENANTIOMERIC MIXTURES Enantiomeric purity tests were made on three analyte pairs, GHG/GhG, LGG/lGG, and YGG/yGG. As the D-enantiomers were added in in-creasing amounts, over the range 1, 3, 5, 10% of the L-enantiomer concentration, the slopes of the correlation lines decreased. Data are shown for 5% "impurity" levels only, FIG. 15.

Judging by the regression coefficient of 0.9998 for the GHG vs GhG plot, there is no significant loss of linearity compared to the reference baseline, meaning that the EP of GhG is equivalent to that of GHG. The explanation given in the Experimental Section for the low percent purity for GhG, as described in the MPS Certificate of Analysis, is apparently vindicated by the results of this spectroscopic method.

Splitting of the YGG/yGG correlation line is consistent with the MPS reported purity level of 97.86% or total impurity of 2.14%.

Noise on the LGG/lGG correlation line conceals whether there is splitting of the line or not. A poor S/N ratio is expected since the CD spectral intensity for LGG is the weakest, FIG. 12,being approximately one-tenth of the band intensities for the other tripeptides.

Enantiomeric excess, defined for example as:

{[Cu(II)(GHG)] - [Cu(II)(GhG)x]} / {[Cu(II)(GHG)] + [Cu(II)(GhG)x]}

is given by the correlation slope for each mixture. Calculated values for spiked GHG solutions are in excellent agreement with the measured values for prepared mixtures, Table 5. Imprecisions based on data from 3-5 repeat measurements, are an improvement by almost a factor of 10 over results obtained from the analyses of binary ephedrine mixtures in which a chemometric analysis method was applied to data at 5 wavelengths. In spite of the splitting of the YGG/yGG and the noise in the LGG/lGG plots, by using best-fit correlation lines, the agreements between calculated and measured EP's are still very good. The method is quantitatively valid over the full range of enantiomeric ratios from 100% L- to 100% D-.

Table 5: Determination of Enantiomeric Purities for Prepared Binary Mixtures of GHG/GhG; LGG/lGG; and YGG/yGG.

% L-form in Prepared Solution	Regression Slope (Enantiomeric Excess)	Regression Coefficient
GHG/GhG		
99	0.9919	0.9998
97	0.9685	0.9998
95	0.9503	0.9999
90	0.8973	0.9999
LGG/lGG		
99	0.9974	0.9951
97	0.9723	0.9957
95	0.9455	0.9946
90	0.8819	0.9932
YGG/yGG		
99	0.9924	0.9996
97	0.9675	0.9996
95	0.9359	0.9996
90	0.8854	0.9995

(b) GGA +"CHEMICAL IMPURITY" LEVELS FOR ALL OTHER DIPEPTIDES.

The question with respect to "chemical impurities" that needs to be addressed is not how great the differences are between the curve for an 8.0mM solution of GGA and curves for the other tripeptides at equimolar concentrations, Figs. 13 and 14, but rather, are the differences sufficient enough to identify and quantitate anonymous chiral "chemical impurities" when these amount to only a few percent of the total composition of a binary mixture? The answer to the question lies in how sensitive the CD detector is in discovering splitting of the correlation line when spectra for "impure" samples are plotted against the spectrum for the primary reference standard.

Spectra were measured for mixtures in which GGA solutions were spiked with small volumes of the other L-tripeptides at levels of 1, 3, 5, and 10%. Data for only the 5% mixtures are plotted in FIG. 16 for the GGX sub-series and in FIG. 17 for GHG, LGG, and YGG. Splittings range from being very small, where they are barely discernable, e.g. for GGI, GGL, and GGF, to extreme, for GGH, GHG, and LGG. Where they are small the best-fit lines, determined by simple linear regression, are seen to

deviate from the unit slope of the reference line. Because of the "absence" of splitting at the lowest concentrations, the plots fail to confirm the presence of GGI, GGL, or GGF at a level of 5% or less, FIG. 16. In general the extreme non-linearity of the split correlations associated with chemical impurities makes it very difficult to determine the amount of impurity.

Briefly recapping the results, the 2-D algorithm has effectively reduced the 1500 spectral data points to one number (the correlation slope), from which EP's can be determined with excellent ac-curacy over the complete range. Recognition that a potential "chemical impurity" is present is elementary for a limited number of cases but its analytical determination is not easily done.

A 3-D Data Reduction Algorithm for Enhancing Selectivity:

The objectives that relate to this sec-ond data reduction algorithm were to discover if the GGA, GGI, GGL, and GGF series can be completely differentiated both qualitatively and quantitatively. The same objectives were achieved when the 3-D algorithm was applied to a series of dipeptides all of which had just one chiral center.

In the 2-D presentations of Figs. 13, 14, 16, and 17, wavelength is an implied variable. For the evolution of the 3-D algorithm, wavelength is the third dimension. Since the experimental parameter measured in CD detection is an absorbance difference, observed signals are positive, negative, and zero. When two CD spectra are plotted against each other, four sign combinations are possible at any wavelength. Repeats of coordinate points, e.g. zero crossover points, can occur at wave-length values that are not adjacent to one another in the spectra. When that occurs 2-D plots "wrap around" and become three dimensional. In retrospect what are observed as 2-D plots are simple projections of the 3-D plots on to the x-y coordinate plane, which explains why some of the plots in Figs.13 and 14 appear to have a 3-D character. The added value of the third dimension is that there should be an increase in the overall analytical selectivity.

The algorithm used for the visual presentation of the three parameter plot was Spinning Plot® which is an integral part of a number of commercially available statistical analyses software packages. The software used for these calculations was JMP 3.1

produced by SAS Institute Inc. Four 3-D plots of wavelength (nm) vs. ellipticity data for GGA vs. ellipticity data for GGA, GGH, GHG, and LGG are shown in FIG. 18. By analogy with the 2-D algorithm procedure GGA plotted against itself is included to provide a baseline for comparison. Front and back quadrants are distinguished by dark and light shading to enhance the 3-D presentation. Discriminations are clearly more evident than they were in Figs. 16 and 17.

Factor Analyses of Spinning Plot® Data:

To derive a quantitative mathematical algorithm, data re-duction was done using a Principal Component Analysis (PCA) procedure on the Spinning Plot® data. Eigenvalues and eigenvectors for the three principal components, P1, P2, and P3, calculated for the GGA/GGH combination plot are given in Table 6. Spatial projections of these same principal components are superimposed on the coordinate axes of FIG.18.

Of the twelve resultant eigenfactors, the one that is most sensitive to variations in the identity of the analyte is P22, highlighted in bold type in Table 6. The 22 tag indicates the entry is in the second row of the second column of the eigenvector matrix. Comparative P22 values for all combinations with GGA are as follows: 0.04519 (vs.GGA); 0.88194 (vs.GGH); 0.13171 (vs.GGI); - 0.01932 (vs.GGL); 0.10312 (vs.GGF); 0.22794 (vs.GHG); 0.89248 (vs.LGG); and 0.57491 (vs.YGG). Standard deviations in P22 determined for data from 3-5 independent repeat measurements are ±0.002, meaning that total analytical selectivity is accomplished. The 3-D algorithm effectively reduced the 1500 original spectral data points to a single discretionary number, P22. The test sets up well in a quality control environment for proving that a chiral substance is or is not a single chemical.

Table 6: Principal Components Calculated for the GGA versus GGH System:

		PC1	PC2	PC3
Eigenvalues		+1.9603	+0.9798	+0.0599
Eigenvectors	nm	+0.19154	+0.97273	-0.13080
	GGA	-0.68522	+0.22794	+0.69175
	GHG	+0.70270	-0.04287	+0.71019

The remaining question is whether the test has the potential to be quantitative. If P22 values were to correlate linearly with the amount of "chemical impurity", then EP's can be determined by difference. Representative plots of P22 vs. percent impurity for solutions of GGA spiked with GGH, GGF, LGG, and YGG are shown in FIG. 19. The plots cease to be linear when the impurity concentration approaches 2.0mM, the concentration of the Cu(II) ion. Differences in the slopes of these lines assist in the identification of the chiral impurity. With the exceptions of GGI and GGL, correlation slopes are greater than two times the ±0.002 SD in the mean for P22 values, which means that impurity levels as little as 1-3% can be measured with confidence provided the impurity is a single chiral substance. Analytical sensitivities are at least ten times more accurate than analogous plots in which maximum ellipticity values measured at a single wavelength are plotted against concentration. Why this is so is easily understood when one sees changes in maximum ellipticity values at a single wavelength over the 1-10% impurity range that are less than the best resolution of ±2.0 mdeg for the CD instrumentation used in this study. The additional accuracy comes from the ability to conveniently include data at 1500 wavelengths.

Where the 2-D method succeeded in providing a means to get accurate values for EP's, the 3-D method provides a way to get a quantitative measure of non-enantiomeric chiral impurities.

Summary and Application of the Method:

By the simple chiral modification of the biuret reagent, combined with two novel data reduction algorithms for the handling of visible CD data, a potentially useful QC

regulatory procedure for peptides, oligopeptides, and proteins has been developed.

A typical procedure begins with the measurement of the visible CD spectrum for the Cu(II) complex of the chromatographically purest available form of the substance being regulated. Data are archived in a computer file on-board the spectrometer and are updated each time the reference material is measured. Spectra for aliquots taken from each newly manufactured product lot are plotted against the standard and successive on-screen visual comparisons are made.

Deviation from a slope of 1.0 in the 2-D test is an instant indication that the purity is less than that of the reference standard. If the value of the regression coefficient indicates an enantiomeric "impurity", the EP is readily calculated from the regression slope. Splitting of the correlation line for the standard reference material gives instant recognition that a "chemical impurity" is present whose identity may be confirmed by the P22 value calculated from the Spinning Plot® algorithm. The percent impurity is calculated from the correlation slope of the P22 vs impurity line.

The two algorithms are complementary in the sense that whereas the 2-D model is capable of measuring EP's with excellent accuracy but only capable of differentiating qualitatively among the eight tripeptides, the 3-D model was capable of quantitatively measuring the compositions of binary mixtures of dissimilar compounds, but incapable of measuring EP's. The latter was not dis-cussed in detail, but is a consequence of the fact the P22 values for an enantiomeric pair are invariant with concentration.

The method is quick, rugged, uses stable inexpensive reagents, requires no specific precautions, and a minimum of technical expertise for a potential operator. Derivatization reactions are instantaneous and data collection is done in a matter of minutes. Spectral data are stored on an on-board computer that is programmed to perform all the mathematical comparisons and quantitative analyses in situ.

All of these advantages point to a very satisfactory and very competitive routine alternative to chromatographic and mass spectrometry methods for the quality control of small peptides.

CHIRAL PROPERTIES OF INSULINS.

Demonstrated successes of circular dichroism (CD) spectropolarimetry detection in enabling direct analyses of biological molecules are often attributable to prior color derivatization reactions that cause the CD activity of the organic material to be shifted into the visible spectral range, far removed from the interferences that affect signal quality in the ultraviolet. Color derivatization alone is a sufficient modification if the analyte is of itself chiral. Reactions where the colored derivatizing reagent is already chiral, e.g. Cu(II)-L-tartrate, introduces even greater spectral variability in that it involves partial ligand exchange with the analyte with the production of diastereoisomers. Specificities related to interactions between the chiral host and chiral analyte ligands introduce a greater level of analytical selectivity.

CD-active derivatizing agents of choice are typically aqueous solutions of first row transition metal ion complexes in which the metal ion contributes the color and the chirality is located on the ligands. Achiral splitting of the ground and excited states of the ligand to metal electronic transitions is the origin of the CD activity. Host ligands contribute a second service to the analyses which is to keep the metal ion from precipitating as an insoluble hydroxide. High pH levels are preferred because the analytes are anionic. Ligands that serve both functions effectively are L-tartrate, (-)-ephedrine (1), and D-histidine. The ligand concentration always exceeds the metal ion concentration by at least a factor of 4.

Factors to consider when choosing a host ligand for the derivatizing agent are the stability of the host complex versus the anticipated stability of the mixed complex formed on ligand exchange, and the complexity of the CD spectrum for the host complex. Copper (II)-D-Histidine and Cu(II)-(-)-ephedrine complexes are the preferred choices for the study of proteins since the thermodynamic stability constants are large (106-1013) and CD spectra for both copper complexes are biphasic, i.e. they consist of more than one band and the bands have opposite signs.

The mechanism for complexing a peptide or protein molecule to first row transition metal ions involves first attachment through the N-atom of the terminal amine followed by ring closure(s) through bonding with N-atoms of the first and succeeding

peptide residues until optimal thermodynamic stability is achieved. Higher substitutions might involve chelation via the O-atom of the terminal acid or amido-group. Substituents on the aminoacid residues generally lie out of the coordinate plane and are factors only in inter- and intraligand interactions, unless a potential Lewis base is present in the side chain.

The full extent of the analytical discriminatory power of visible range CD detection data for copper-peptide complexes was more than adequately demonstrated in the total differentiation among the members of a series of eight one-chiral-center tripeptides and a series of ten dipep-tides, two of which were enantiomers. Mathematical algorithms were derived that enhanced the selectivity and expressed the total differentiations in numerical terms. Departures from these characteristic numerical values were used to mathematically determine the amounts of related chemical impurities, and to measure EP's with an unprecedented accuracy of ±0.25% over the whole mixture range from 52-99%.

In advancing the study from these relatively simple peptide structures to protein molecules and the many subtleties and inherent complexities that they present, the question is whether this simple bonding mechanism, coupled with specific interactions within the 3-D architecture surrounding the extended coordination sphere of the tetragonal Cu(II) ion, are sufficient to differentiate among protein analogs that are structurally so very similar. In the first investigation of this question, a study was made on four commercially available insulin products, human, human Lyspro, porcine, and bovine, and on the separated A and B chains of bovine insulin.

Experimental:

Reagents: Reagent grade D-histidine.HCl (Sigma) was the preferred host ligand. Reagent grade CuSO4 .5H20 was a product of Fisher Scientific.

Insulin forms used for the study were taken from different commercially available master lots obtained from three sources, Lilly, Novo Nordisk, and Sigma Chemical. Altogether 38 lots were accessed, though not all were quantitated. Some lots were in the Zn-crystalline form (Lilly, Novo), others were Zn-free (Novo). Origins for human

insulins were E. coli (Sigma), yeast (Sigma), and recombinant DNA expressions (Lilly, Novo). Bovine insulins (Sigma, Lilly) were described as Beef Purified or from bovine pancreas. Lilly and Sigma were the sources for Pork Purified Zn-porcine insulin crystals. Different lots of the Lyspro variant of human insulin were obtained from Lilly. Bovine insulin A and B chains were both Sigma products.

Solution preparations:

Stock solutions of the Cu(II)-(D-histidine) host complex were prepared in such a way that [Cu2+] = 0.020M, and [D-histidine] = 0.080M. The pH was adjusted to 13.0 with added NaOH. KI at a concentration of 0.30M is added as a stabilizer. Stock solutions are stable for several weeks, and usually used until exhausted. Working solutions for the host complex were prepared by diluting stock solutions by a factor of 10 using 0.10M NaOH. A typical derivatizing agent, therefore, is a solution that is 2.0mM in Cu(II), 8.0mM in D-histidine, 0.1M in NaOH, and 0.03M in KI.

Appropriately weighed insulin samples were added directly to 15.0mL aliquots of the host complex reagent to obtain an analyte concentration of 150 M, a concentration that is far below the amount that would fully exchange with the D-histidine at 8.0mM. An insulin concentration of 150 M was chosen because, from the results of a preliminary titration study, that particular concentration was observed to produce the greatest spectral differences between the host and the different mixed complexes.

Measurements:

Spectra were measured using a Jasco 500-A automatic recording spectropolarimeter coupled to an IBM-compatible PC through a Jasco IF-500 II serial interface and data processing software. Experimental parameters were: wavelength range 400-700nm; sensitivity 20 mdeg/cm; time constant 0.25s; scan rate 200 nm/min; single scan; pathlength 5.0cm; temperature ambient. Calibration of the day to day reproduciblity of the system was done by measuring the CD spectrum for the host Cu(II)-D-histidine solution. Standard deviations from the mean for the day to day reproducibility of the maximum ellipticities measured at wavelengths 487nm and 682nm were 7.42 ±0.07

mdeg and -214 ±0.60 mdeg respectively.

Insulin additions were made immediately prior to analysis. Spectral measurements were timed to start at 3-5 minutes after mixing. This precaution was taken to preclude any spectral changes that might occur during conceivable changes in the 3-D structure of the protein with time. With no evidence to the contrary, the assumption was made that in solutions at pH 13.0 the dissolved insulins are monomeric.

Results and Discussion.

With the D-histidine in excess over the insulin by a factor of 500, the law of mass action lies heavily in favor of preferentially binding D-histidine to the Cu(II) ion. Ligand exchange, therefore, is much less than stoichiometric, so the final product of the exchange is a mixture of two CD-active complexes; the original host and the mixed Cu(II)-{D-histidine-insulin} complex. The presence or absence of Zn2+ ion in the preparation is not a factor either because the ratio of the [Cu2+] to [Zn2+] in test solutions is also in excess of 500/1. Since two very stable complexes are formed in mutual equilibrium, the correlation of CD signal with analyte concentration will be non-linear, making it a requirement that the same total concentration be maintained for all of the analytes. Because of the strong molecular similarities of the four insulin types, it is assumed that the stability constants of all the mixed complexes will be identical and not a factor to consider in explaining any differences that may be seen in the CD spectra after addition of the analytes.

Representative spectra for the host complex and the mixed complexes with human, human (Lyspro), porcine, and bovine insulins are shown in FIG. 20. Spectra for the human and porcine insulins (all sources and all lots), which differ only in the identity of the B31 residue, appear to be equivalent, which is not unexpected. Differences do however exist between these two and the spectra for bovine and human (Lyspro).

Repetitive measurements were made over a period of several months on samples taken from: ten human insulin master lots obtained from Lilly (n=31), Novo (n=22), and Sigma (n=2); three porcine lots from Lilly (n=24) and Sigma (n=7); four bovine lots from Sigma (n=7) and Lilly (n=8); and three Lyspro lots from Lilly (n= 22). Several new preparations of the host complex were required over the duration of the study and spectra

were found to be very reproducible, in ac-cord with the mean SD's reported in the solutions preparation section.

In evaluating robustness, comparisons were made among lots from a single manufacturer, between manufacturers, and on consecutive preparations from the same lot taken as a function of time, often as much as one month apart. Spectral reproducibilities for the mixed complexes, based on measured signals at the maximum wavelengths, were better than 2%, testifying to the rugged-ness of the employed technique.

Reliance on the goodness of fit when spectra for the same insulins from different manufacturers are superimposed is too subjective a test to verify that all samples are equivalent and that all conform with the enantiomeric properties of a single standard reference material. A more quantitative algorithm is essential to validation in a quality control context.

A 2-D Data Reduction Algorithm for Enhancing Selectivity:

As a first attempt at deriving a quantitative algorithm, spectral ellipticity data (n=1500 points) for D-histidine were plotted against the spectral data for the insulins, FIG. 21. The evidence that the insulins are not chemically equivalent is quite clear. The non-linear nature of these cross correlations makes quantitation difficult.

As a second attempt at quantitation, spectra for pairs of insulin samples are plotted against one another. If the materials are chemically identical, the correlation will be a perfect straight line with a slope and a correlation coefficient of 1.0. For materials that are the same but of unequal purity, the correlation coefficient will again be 1.0 but depending upon the chirality of the "impurity", the slope of the straight line will be either greater than or less than 1.0. Correlation lines are neither straight nor have a slope of 1.0 for materials that are not the same, as seen in FIG. 22 where human insulin data are plotted against data for human Lyspro, porcine, and bovine forms.

Results in Figs. 21 and 22 provide compelling evidence that there are real chiral differences between human insulin and the Lyspro and bovine insulins when complexed to Cu(II) ion as would be expected from the zero order spectra of FIG. 20. The magnitude

of the departure from linearity in the case of human vs. Lyspro is very surprising considering that the only structural difference between them is the mutual exchange of the proline-B29 and lysine-B30 residues. Some evidence that the plot of human vs. porcine is not exactly collinear is observable at ellipticity values less than - 100mdeg. This "departure" accounts for a 2% increase in the correlation slope to 1.021, FIG. 22, but is probably not sufficient proof that differentiation has been achieved. We shall return to this point later. Considering that the only difference in the peptide sequences for human and porcine insulins is in the terminal B-31 residue, and that it is physically remote from the residues at the amine terminus that are involved in bonding to the Cu(II) ion, perturbations from the human structure would be expected to be negligibly small, and very difficult to detect experimentally.

Realistically speaking, the mutual exchange of proline with lysine in the human/ human Lyspro comparison may be representative of the ultimate limit to the analytical selectivity of this assay in the QC validation of insulins. The confidence level in differentiation after the 2-D test is, at best, three out of four. Compared with the simple superposition of spectra, FIG. 20, decisions on whether substances are equivalent or not are better founded from the spectral data correlation plots since they are based on simple numerical information.

A 3-D Data Reduction Algorithm for Enhancing Selectivity:

In the 2-D presentations of Figs. 21 and 22, wavelength is an implied variable. For the evolution of a 3-D algorithm, wavelength is the third dimension. Spin Plots® were derived for every spectrum for every sample using wave-length (x-axis), ellipticity data for the D-histidine host complex (y-axis), and ellipticity data for the mixed complexes (z-axis).

Since the experimental parameter measured in CD detection is an absorbance difference, ob-served signals are positive, negative, or zero. When two CD spectra are plotted against each other, four sign combinations are possible at every wavelength. Repeats of (y,z) coordinate points can occur at wavelength values (x1...xn) that are not adjacent to one another in the spectra. When that happens 2-D plots "wrap around" and

become three dimensional. In retrospect what are observed as 2-D plots in Figs. 21 and 22 are projections of the 3-D plots on to the x-y coordinate plane. 3-D plots are shown for human Lyspro and bovine insulins in FIGS. 23 (A) and (B) respectively. FIG. 23 can be transformed into FIG. 21 by projecting the 3-D plot vertically on to the (y-z) plane. The added value of the third dimension is that with an additional variable, there should be an increase in the overall analytical selectivity.

Data Reduction by Principal Component Analyses:

In order to make quantitative comparisons among the figures for each insulin, data reductions were made using principal component analyses (PCA) (or factor analyses) of the full range of spectral data. Most standard statistical software packages are equipped to handle PCA algorithms routinely. PCA results are expressed in matrix form as eigenvalues and eigenvectors for each of the principal components for each of the analytes. An example is given in Table 7 for a Lilly human insulin sample.

Viewing a spin plot is not an essential step in the PCA. Plots are included here only as pictorial representations of what might conceivably become "fingerprints" with which to identify the insulins by type. Principal component axes, P1, P2, and P3, added to FIG. 23, are a very useful aid in effecting distinctions. Notice the different lengths (eigenvalues) for P1 and the different spatial orientations (eigenvectors) for P2 between the figures (A) and (B).

Spin plots and PCA calculations were made for every spectrum for every individual sample that was tested. Factors in Table 7 that are the most sensitive to the identities of the insulins are the four values given in bold type in rows 2 and 3, subsequently referred to as P22, P23, P32, and P33 respectively. Correlating variations in any one or all of these numbers with the analytes in question is a prospective route to achieving 100% analyte differentiation. How good the differentiation really is can be judged by comparing the factor values for all insulin types, Table 8. Numbers given are the statistical means from repeated measurements listed by manufacturer, by numbers of lots as-sayed, and as the overall means. Standard deviations (SD) are included to assist in the decision making.

Table 7: Data Reduction by Principal Component Analysis of a Spinning Plot Presentation of CD Data for the Host and Mixed Copper Complexes of Human Insulin.

Principal Components	P1	P2	P3
Eigenvalues	2.4993	0.4813	0.1094
Eigenvectors			
nm	-0.6270	0.1122	0.7709
D-Histidine	0.5290	0.7878	0.3155
D-Hist/Human Insulin	0.5719	-0.6056	0.5532

Table 8: Comparisons of Mean Values for P22, P23, P32, and P33 Eigenvectors for D-Histidine/Insulin Complexes Among Insulin Types and Manufacturers.

Type	Source (repeats)	No. of Lots	Mean			
			P22	P23	P32	P33
Human	Lilly (31)	5	0.78780 (0.0059)	-0.60256 (0.0056)	0.31553 (0.0054)	0.55327 (0.0060)
	Sigma (2)	1	0.79896 (0.0018)	-0.58383 (0.0014)	0.28271 (0.0020)	0.57109 (0.0011)
	Novo (22)	4	0.78735 (0.0079)	-0.60805 (0.0076)	0.31792 (0.0071)	0.55251 (0.0068)
	ALL (55)	10	0.78851 (0.0056)	-0.60425 (0.0055)	0.31277 (0.0056)	0.55442 (0.0054)
Porcine	Sigma (7)	1	0.79023 (0.0072)	-0.59974 (0.0067)	0.30852 (0.007)	0.55698 (0.0066)

	Lilly (24)	2	0.79438 (0.0058)	-0.59354 (0.006)	0.29574 (0.0056)	0.56363
	ALL (31)	3	0.79510 (0.0083)	-0.59267 (0.0078)	0.29464 (0.008)	0.56464
Human Lyspro	Lilly (22)	3	0.73437 (0.0077)	-0.67293 (0.0088)	0.42275 (0.0084)	0.48814 (0.0081)
Bovine	Sigma, Lilly (15)	4	0.82594 (0.0098)	-0.52442 (0.0110)	0.16287 (0.0099)	0.63266 (0.0101)
Bovine Chain A	Sigma	1	0.02123	-0.69856	0.77783	0.46695
Bovine Chain B	Sigma	1	-0.13698	0.91557	0.67531	0.37435
A/B Equimixture	Sigma	1	0.00163	0.71310	0.74058	0.47798
A/B Spectral	Sigma	1	0.00211	0.71226	0.74093	0.47851
Average						

Given that SD's are never larger than 0.010, the mathematical and statistical evidence confirm that all of the human insulin lots from Lilly and from Novo that were sampled are equivalent. This is in excellent agreement with the conclusion reached from there being an almost perfect linear plot in FIG. 22. Data for the Sigma human samples are apparently different but the number of samples measured is not statistically significant. Evidence that Lyspro and bovine insulins are distinguish-able from each other and from human and porcine insulins is indisputable. The advantage of this algorithm over the non-linear correlation plots of FIG. 22 is that the bases for the differentiation judgments are now four numerical values that might ultimately turn out to be unique to that protein. Differences in the magnitudes of the four factors for porcine and bovine insulins are clearly large enough to quantitate the proportions of each in bovine-porcine mixtures. Even better accuracies would be obtained if the concentrations in the working solutions were increased.

In contrast with the other analogs, calculated means for P22, P23, P32, P33 for human (n=55) and porcine (n=31) insulins are quantitatively very similar. Because of a slight overlap at the extremes of the (SD) ranges for the two analytes, it is a reasonably legitimate argument to say that human and porcine insulins are indistinguishable. But based upon the small but clear departure from perfect linearity in the 2-D spectral correlation plot for human and porcine in FIG.22, there is reason to suspect that the means are statistically different. The latter argument is substantiated by repeating the Spin Plot and PCA data treatment with one modification, D-histidine is replaced by human insulin as the common reference spectrum. If it happens that the P22, P23, P32, and P33 values, calculated for (wavelength vs. human vs. human), are not identical with the values calculated for (wavelength vs. human vs. porcine), then the chemical compounds are not the same, Table 9. The evidence is that the analytes are not identical in terms of their chiral properties and are therefore not the same molecule. A successful distinction between two proteins that differ by only one aminoacid residue is something that could not have been deduced from either the subjective overlap of spectra or from Beers' Law plots derived using data at only a single wavelength

Table 9: Comparisons of the Mean Values for the P22, P23, P32, andP33 Eigenvectors when Human Insulin is the Reference Material Replacing D-Histidine in Table 8 **.

Type	Mean P22	Mean P23	Mean P32	Mean P33
Human vs. Human	0.40079	- 0.40079	0.70711	- 0.70711
Human vs. Porcine	0.44098	0.36046	- 0.68386	0.72806
Human vs. Lyspro	0.14921	0.60849	0.79096	- 0.54648
Human vs. Bovine	0.82594	- 0.52442	0.16287	0.63266

**Standard deviations are on the same orders of magnitude as those in Table 8.

The amplification of the differences in the values between Tables 8 and 9 happens for the following reason. The D-histidine ligand is the dominant contributor to the CD spectra in the first option, and is factored out in the second. Discrimination between human and porcine insulins is now eminently obvious, and is especially manifest by the sign reversals for P32 and P33. Effects of factoring out D-histidine in the analyses of Lyspro and bovine insulins are included in Table 9 for completeness.

The "chiral recognition" of changes in a peptide sequence at positions that are so far removed from the active binding site, has to be among the ultimate achievements in selectivity by analytical spectroscopy. If substantiated with further work, future applications of the method to QC of peptide drug forms are potentially endless.

If the accepted mechanism that a protein complexes to a metal ion through involvement of the residues at the amine terminal is in fact correct, then it is difficult to completely understand the causes behind the very significant CD changes observed for the human, Lyspro, bovine, and porcine insulins. In the first place, all four have the same initial aminoacid sequences in both the A- and B-chains. Secondly, the sequence changes that bestow the different properties to each insulin are at locations remote from the amine ends. Why spectral differences occur has to be accounted for in other ways: (a) the 3-D architecture of the protein in the extended coordination sphere of the Cu(II) ion is significantly altered on binding; (b) additional substitution(s) at the axial coordinate positions normally occupied by hydroxide ion at pH 13; (c) involvement of the amine

ends of the A-and B-chains that differ from one insulin to another; and (d) ligand exchange has not occurred at all, in which case the spectral changes are attributable to organized hydrophobic interactions between the coordinated D-histidine ligands and the proteins only.

Bovine A- and B-chains:

The A- and B-chains of bovine insulin, were examined separately. Their importance to the question is that their initial sequences are different, Gly.Ile.Val.Glu- for the A-chain vs. Phe.Val. Asp.Glu- for the B-chain, but they are common to all four insulins. Evidently both chains are capable of exchanging with D-histidine, FIG. 24. The CD spectra are completely different from each other and from the spectrum for the intact insulin, as are the calculated P2 aand P3 values (Table 8). The results indicate that a sensitivity to the initial sequence exists which suggests that the exchange mechanism is correct.

CD spectra that are measured for a prepared equimolar mixture of the A- and B-chains are superimposable on the spectrum that is the calculated mean of the spectra for the separate chains, FIG.24. This is an indication that the two chains exchange independently, at least when they are separate. It does not prove however that both are involved when the insulin molecule exchanges. If they are, then multiple substitution is an important factor to consider when accounting for analytical selectivity. The practical result of this experiment is that another step is taken towards achieving the ultimate analytical specificity and augurs well for future studies into applications of the method to QC validations of peptide and protein drug forms.

Summary:

A method to validate the chiral properties (aminoacid sequences and enantiomeric substitutions) and the chemical purities of proteins, and peptides, for the purposes of QC of commercial drug products has been described. The method is quick, very simple to perform, and is proven to be experimentally rugged. Mathematical algorithms are introduced by which full spectral data can be reduced to a few discretionary numbers that have the potential to become a "characteristic" values for every peptide species, once it

has been tested for a library of related compounds. If vindicated, the level of analytical selectivity among chiral molecules is unparalleled.

While the invention has been described with a certain degree of particularity, it is manifest that many changes may be made in the details of construction without departing from the spirit and scope of this disclosure. It is understood that the invention is not limited to the embodiment set forth herein for purposes of exemplification, but is to be limited only by the scope of the attached claim or claims, including the full range of equivalency to which each element thereof is entitled.

CLAIMS

What is claimed is:

1. A chemical reagent solution, comprising:
a mixture of $Cu(SO_4)\cdot 5H_2O$;
D-histidine; and,
NaOH.

2. The reagent solution of claim 1 wherein KI is added as a stabilizer.

3. The reagent solution of claim 1 wherein:
the Cu^{2+} concentration is approximately 2.0mM;
the D-histidine concentration is approximately 8.0 mM; and;
the NaOH concentration is approximately 0.1M.

4. The reagent solution of claim 3 wherein the pH is greater than 12.

5. The reagent solution of claim 4 wherein the pH is approximately 13.

6. The reagent solution of claim 3 further including KI at a concentration of approximately 3.0 mM.

7. A process for differentiating a peptide assay, comprising:
(a) obtaining a volume of reagent solution including a Cu(II) metal ion and a chiral host ligand;
(b) adding an aliquot of analyte peptide such that exchange occurs between said host chiral ligand and said analyte peptide to form a metal-peptide complex;
(c) obtaining a CD spectrum on said metal-peptide complex.

8. The process of claim 7 including the further steps of:

(d) obtaining the CD spectrum of a standard peptide wherein said standard peptide is the pure form of said analyte peptide;

(e) comparing said CD spectrum of said analyte peptide with the CD spectrum of said pure peptide in order to determine the purity of said analyte peptide.

9. The process of claim 7 wherein a analyte protein is substituted for said analyte peptide.

10. The process of claim 7 wherein the concentration of said analyte peptide does not exceed the concentration of said Cu(II) ion.

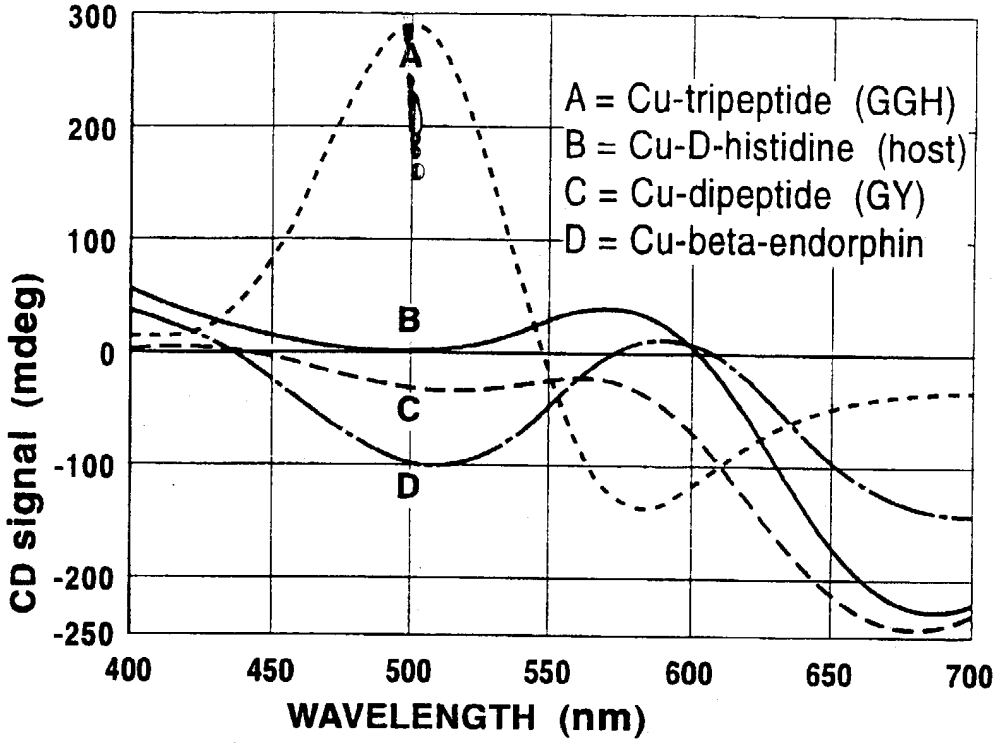
A = Cu-tripeptide (GGH)
B = Cu-D-histidine (host)
C = Cu-dipeptide (GY)
D = Cu-beta-endorphin
CD signal (mdeg)
WAVELENGTH (nm)
300
200
100
0
-100
-200
-250
400
450
500
550
600
650
700

Fig. 1A

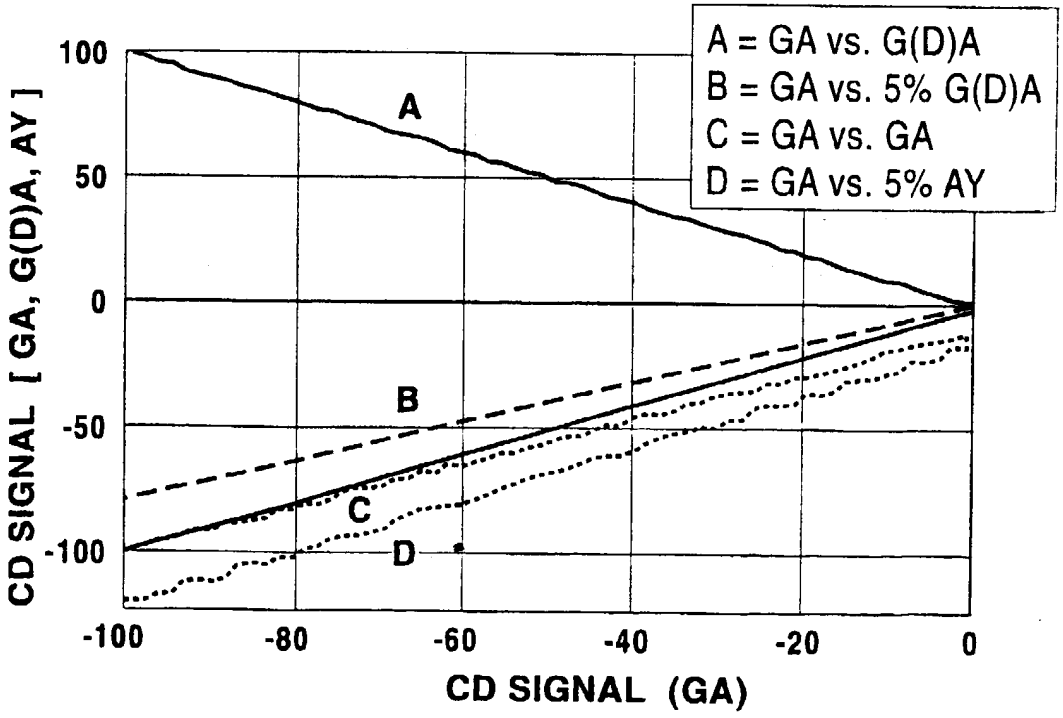
A = GA vs. G(D)A
B = GA vs. 5% G(D)A
C = GA vs. GA
D = GA vs. 5% AY
CD SIGNAL [GA, G(D)A, AY]
CD SIGNAL (GA)
100
50
0
-50
-100
-100
-80
-60
-40
-20
0

Fig. 1B

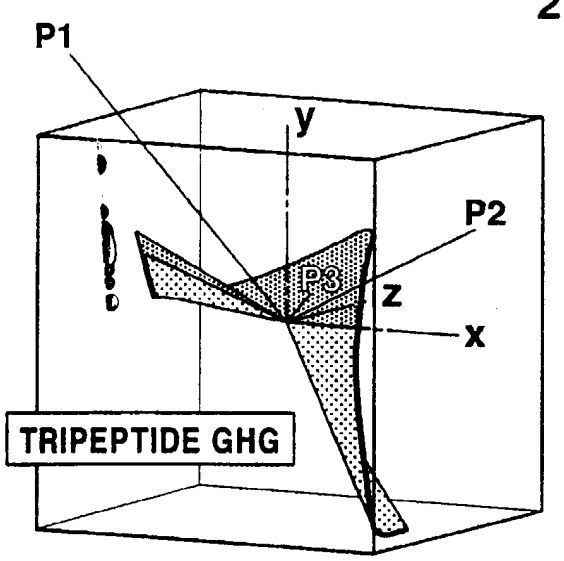

Fig. 2A

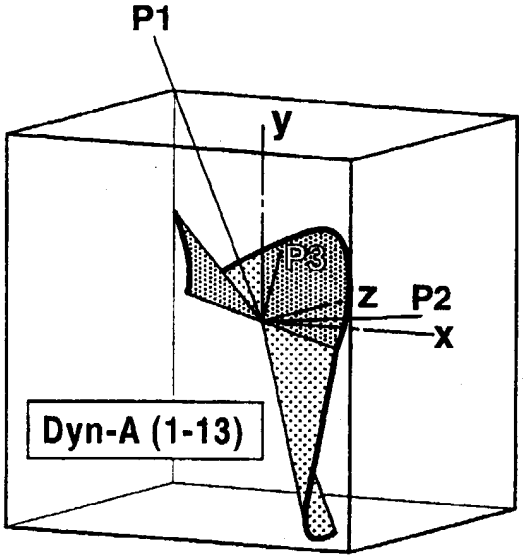

Fig. 2B

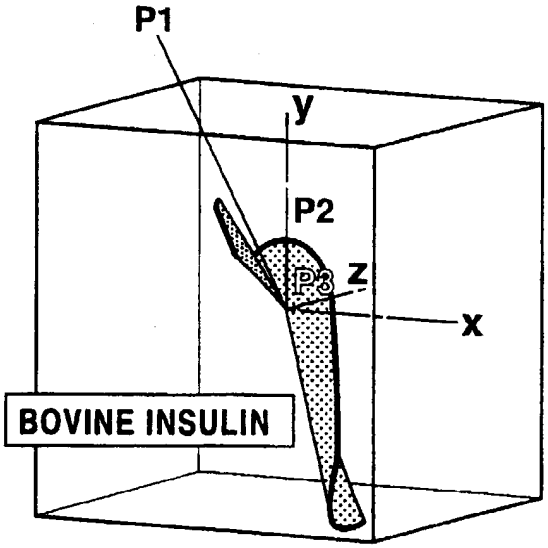

Fig. 2C

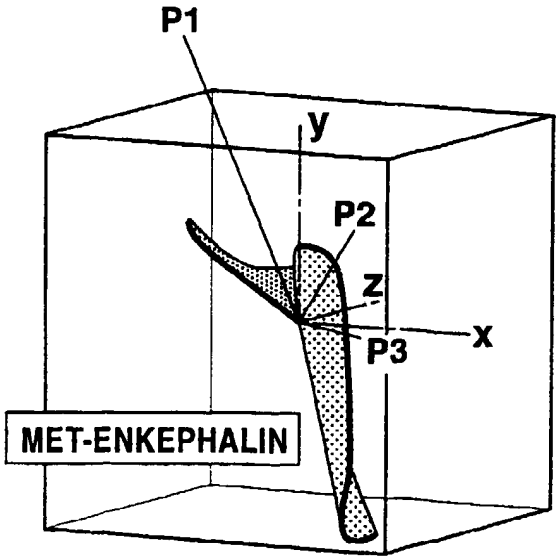

Fig. 2D

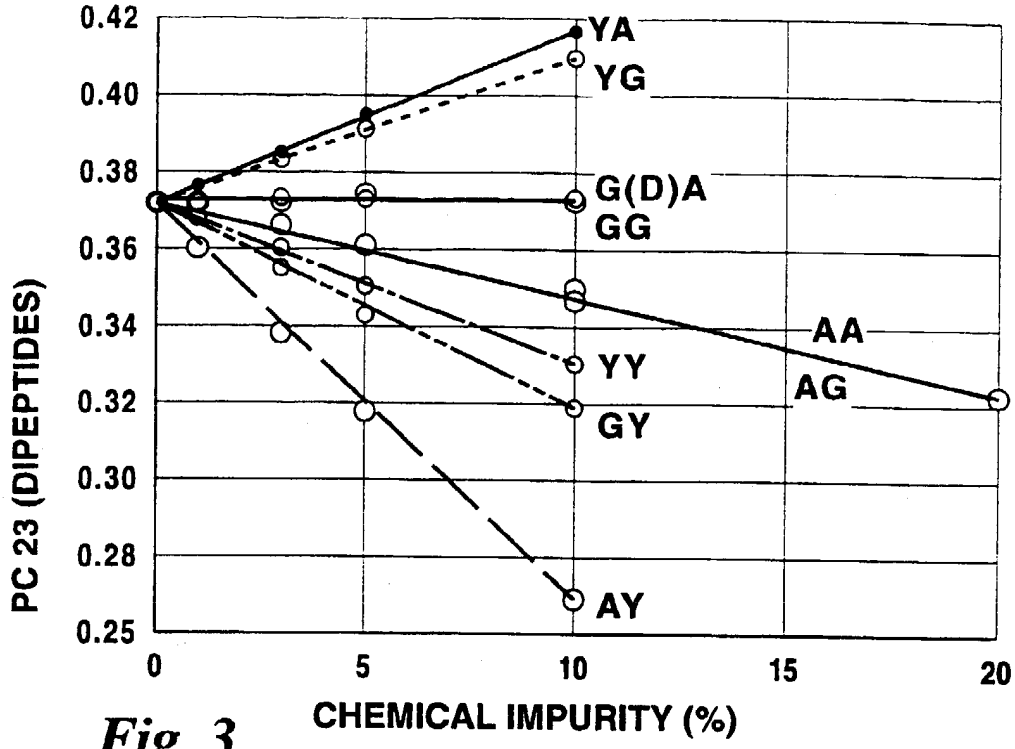
0.42
0.40
0.38
0.36
0.34
0.32
0.30
0.28
0.25
PC 23 (DIPEPTIDES)
YA
YG
G(D)A
GG
AA
AG
YY
GY
AY
0
5
10
15
20
CHEMICAL IMPURITY (%)

Fig. 3

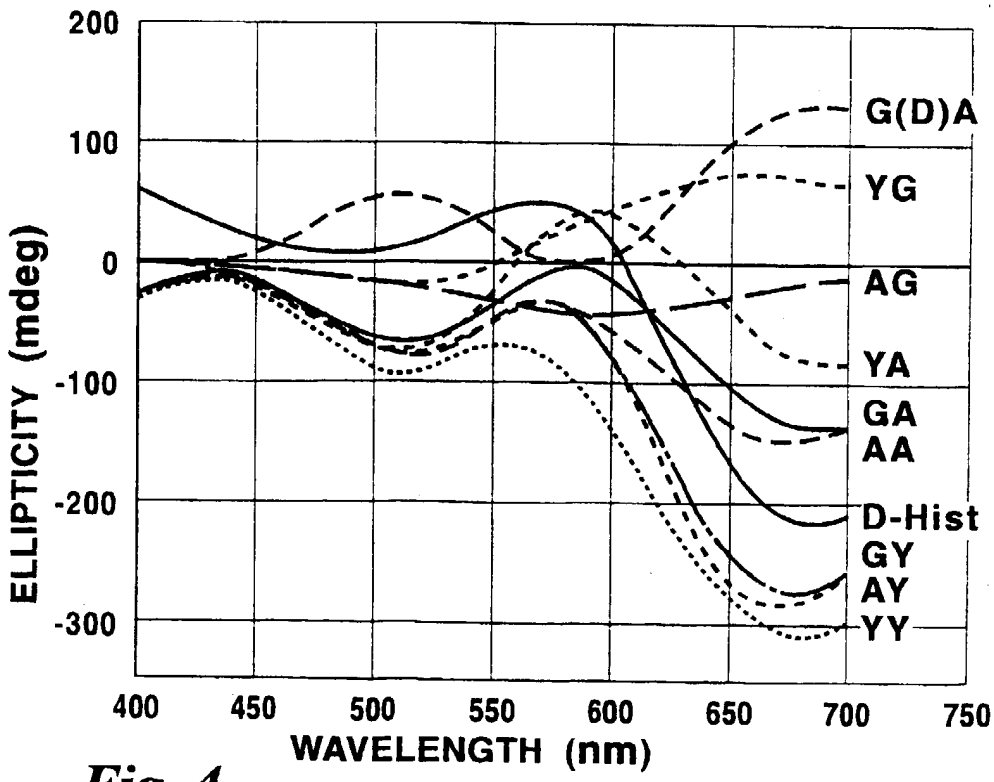
200
100
0
-100
-200
-300
ELLIPTICITY (mdeg)
G(D)A
YG
AG
YA
GA
AA
D-Hist
GY
AY
YY
400
450
500
550
600
650
700
750
WAVELENGTH (nm)

Fig. 4

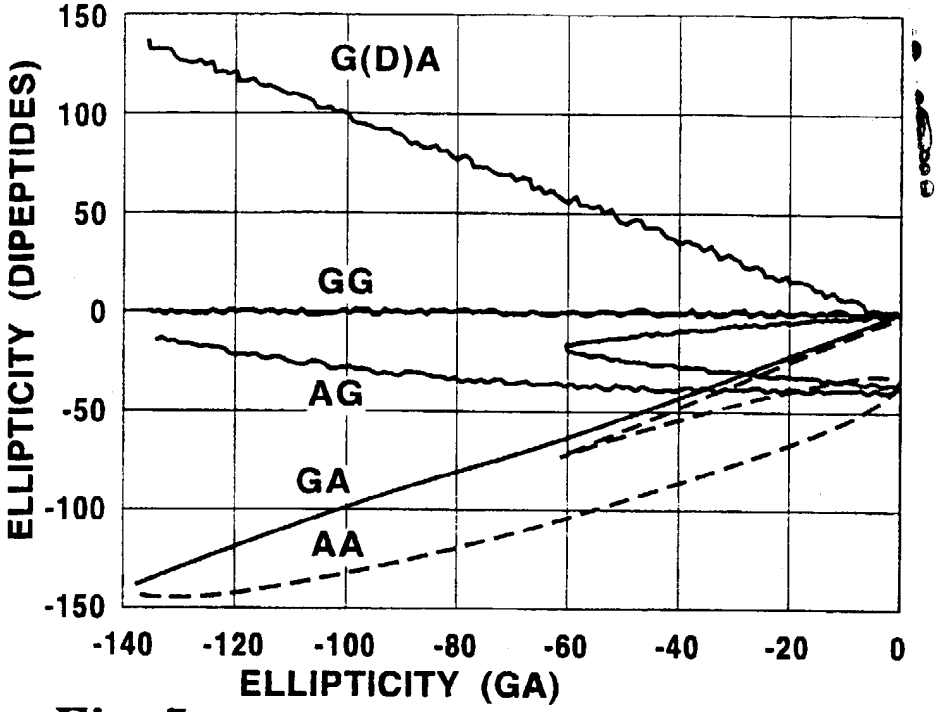
150
100
50
0
-50
-100
-150
ELLIPTICITY (DIPEPTIDES)
G(D)A
GG
AG
GA
AA
-140
-120
-100
-80
-60
-40
-20
0
ELLIPTICITY (GA)

Fig. 5

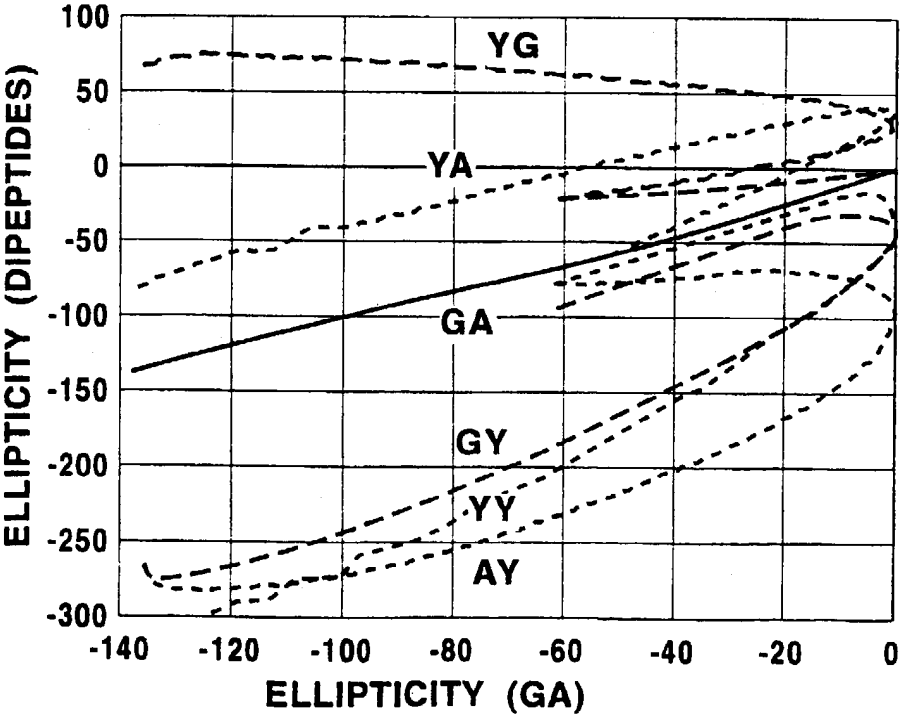
100
50
0
-50
-100
-150
-200
-250
-300
ELLIPTICITY (DIPEPTIDES)
YG
YA
GA
GY
YY
AY
-140
-120
-100
-80
-60
-40
-20
0
ELLIPTICITY (GA)

Fig. 6

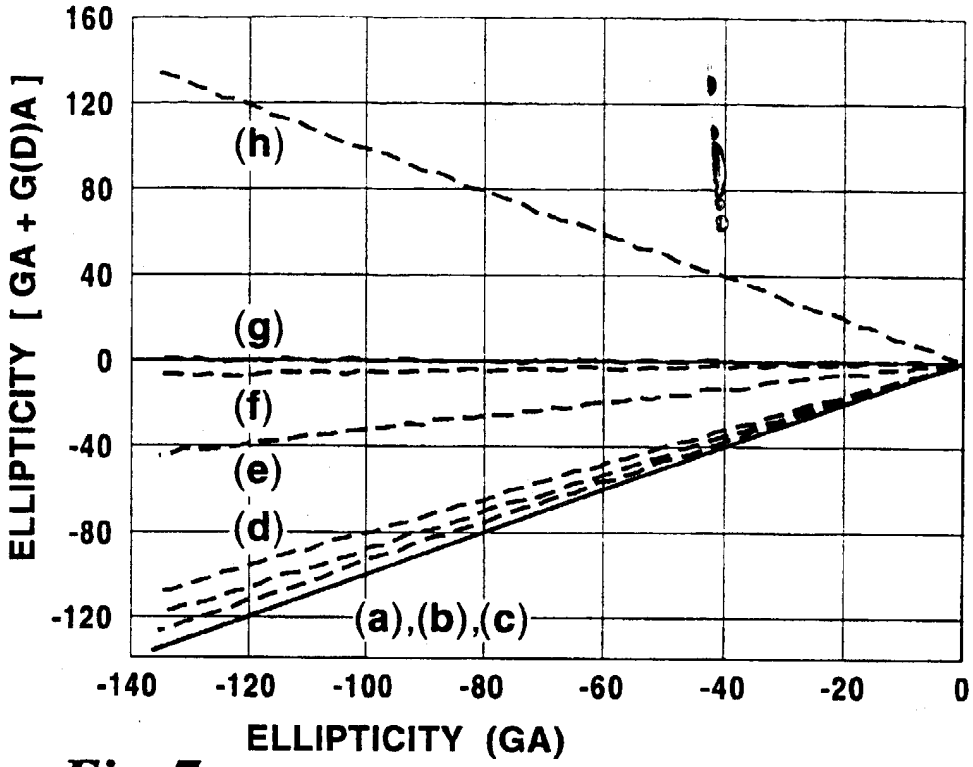
160
120
80
40
0
-40
-80
-120
ELLIPTICITY [GA + G(D)A]
(h)
(g)
(f)
(e)
(d)
(a),(b),(c)
-140
-120
-100
-80
-60
-40
-20
0
ELLIPTICITY (GA)

Fig. 7

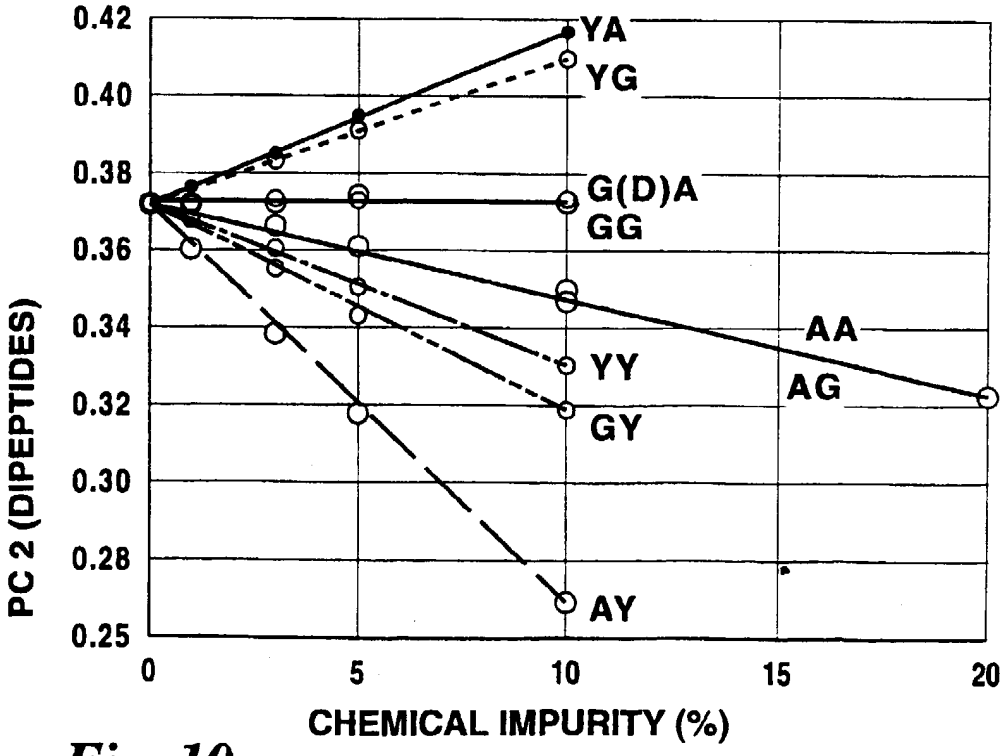
0.42
0.40
0.38
0.36
0.34
0.32
0.30
0.28
0.25
PC 2 (DIPEPTIDES)
YA
YG
G(D)A
GG
AA
AG
YY
GY
AY
0
5
10
15
20
CHEMICAL IMPURITY (%)

Fig. 10

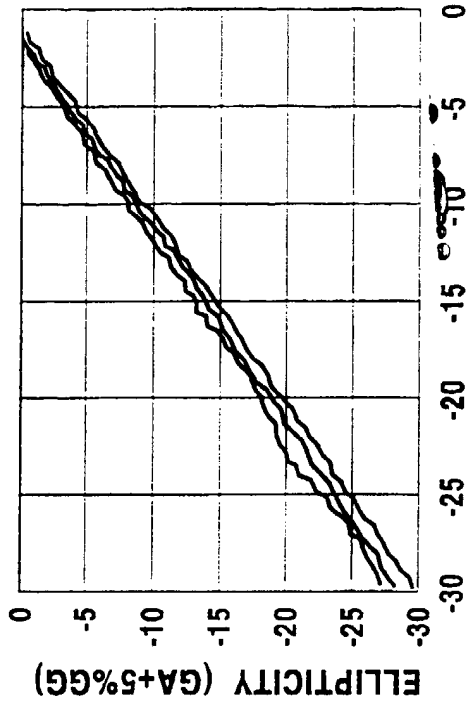

Fig. 8B

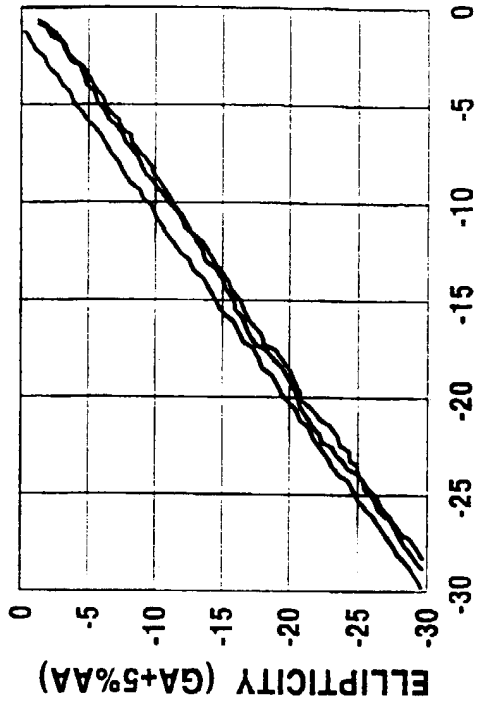

Fig. 8D

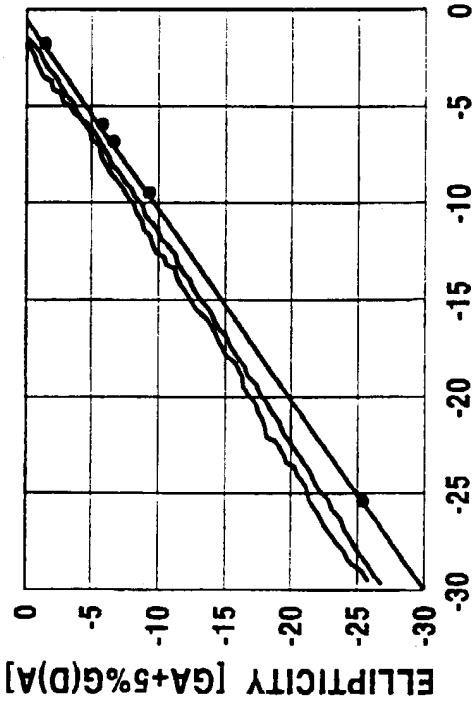

Fig. 8A

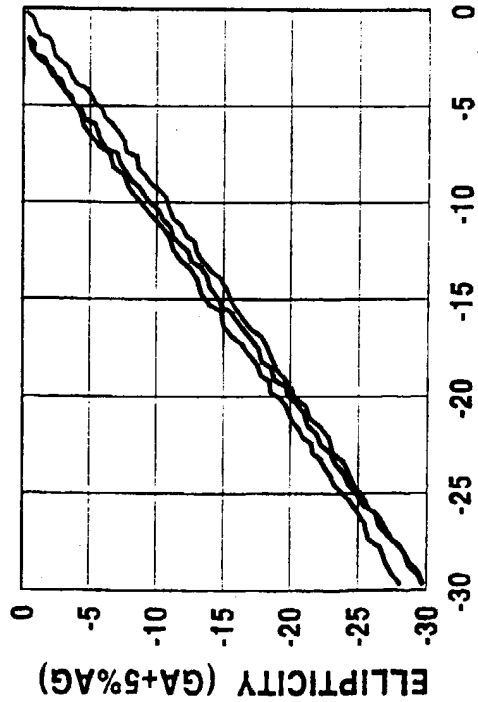

Fig. 8C

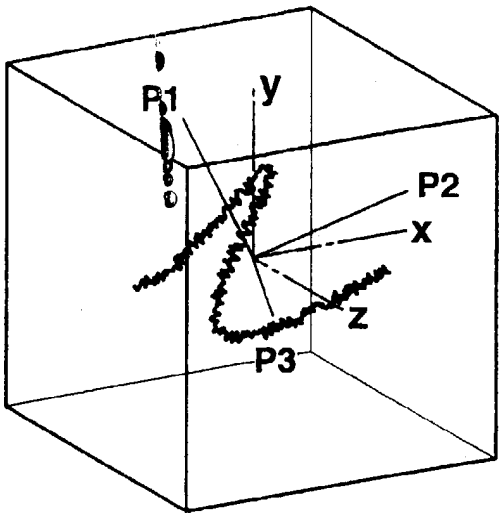

Fig. 9A

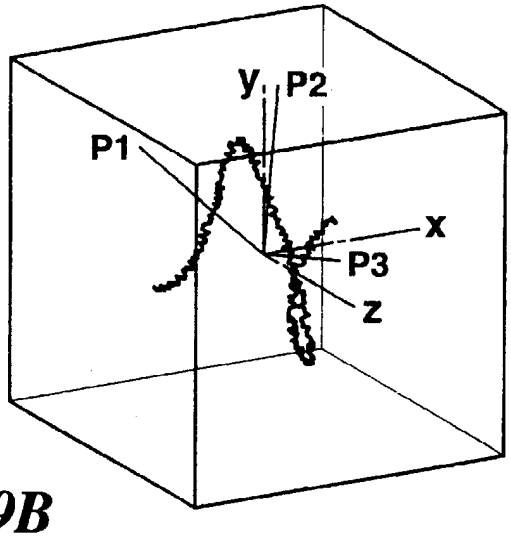

Fig. 9B

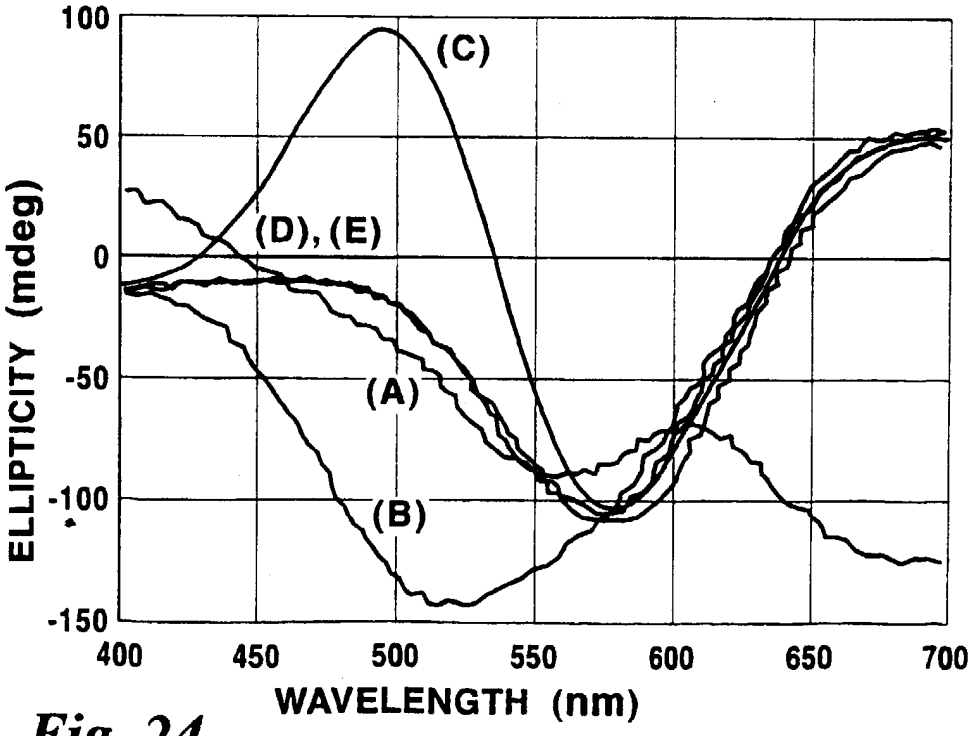

Fig. 24

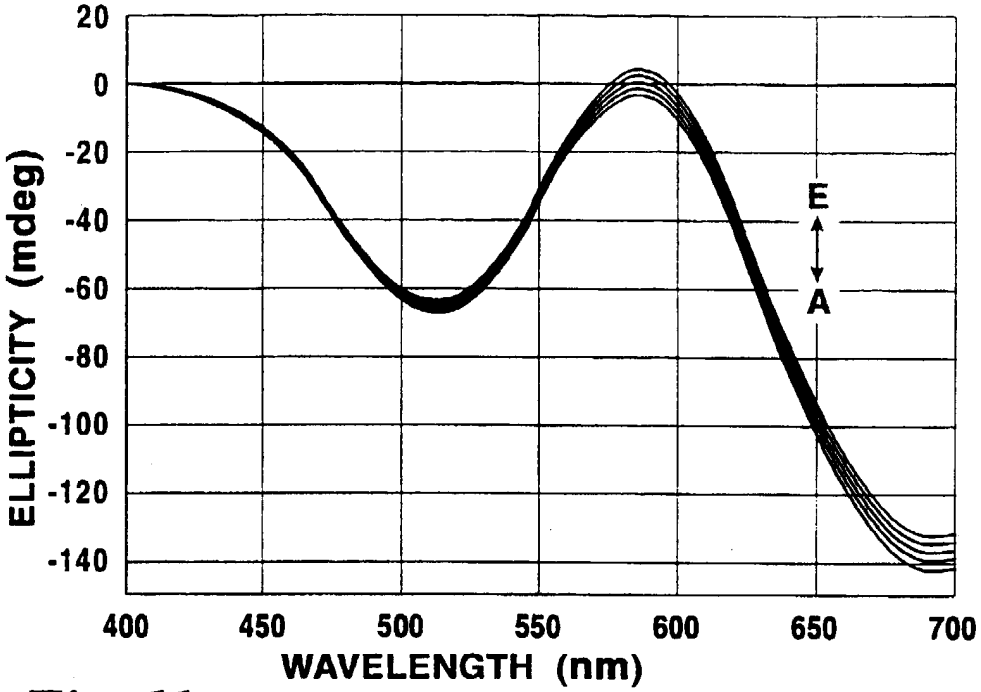
20
0
-20
-40
-60
-80
-100
-120
-140
ELLIPTICITY (mdeg)
400
450
500
550
600
650
700
WAVELENGTH (nm)
E
A

Fig. 11

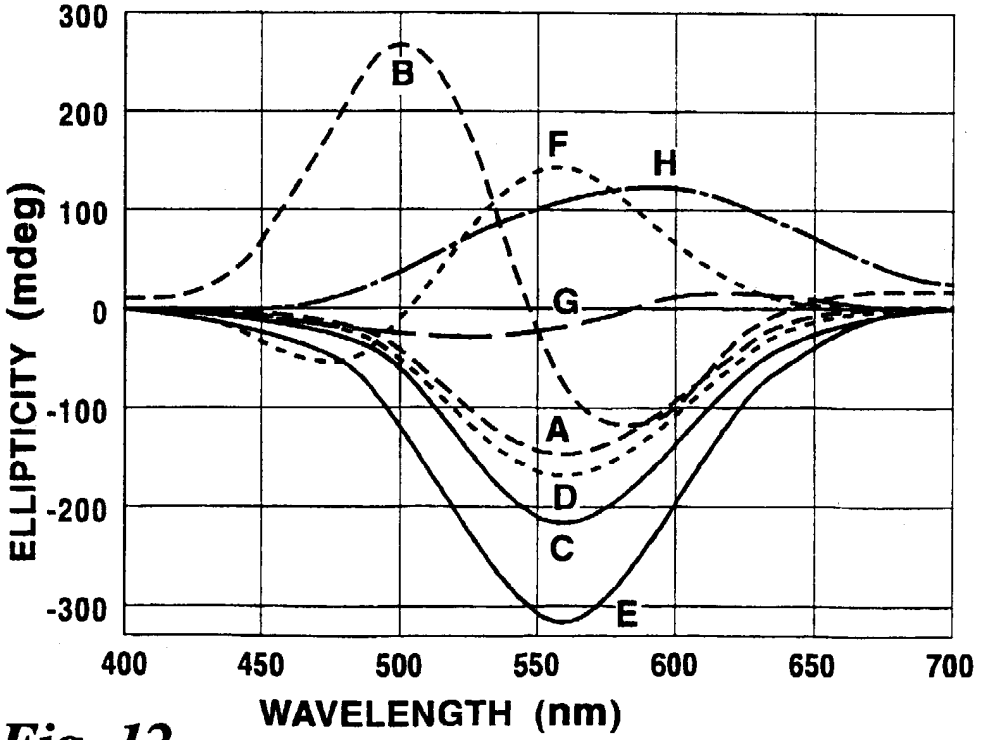
300
200
100
0
-100
-200
-300
ELLIPTICITY (mdeg)
400
450
500
550
600
650
700
WAVELENGTH (nm)
B
F
H
G
A
D
C
E

Fig. 12

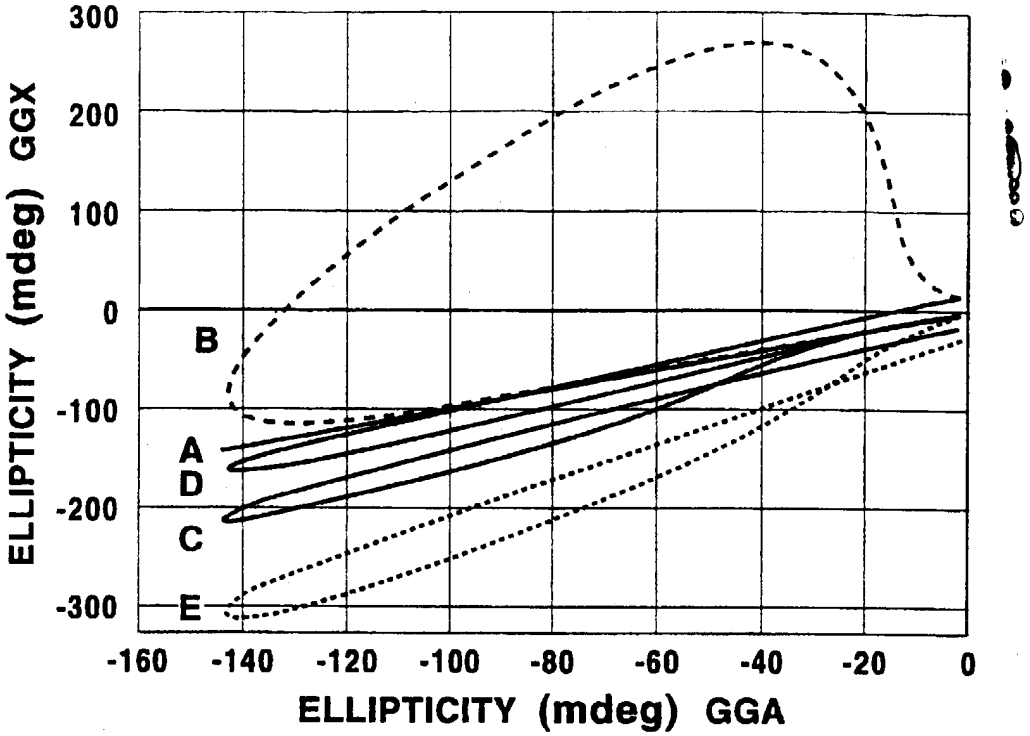
300
200
100
0
-100
-200
-300
ELLIPTICITY (mdeg) GGX
B
A
D
C
E
-160
-140
-120
-100
-80
-60
-40
-20
0
ELLIPTICITY (mdeg) GGA

Fig. 13

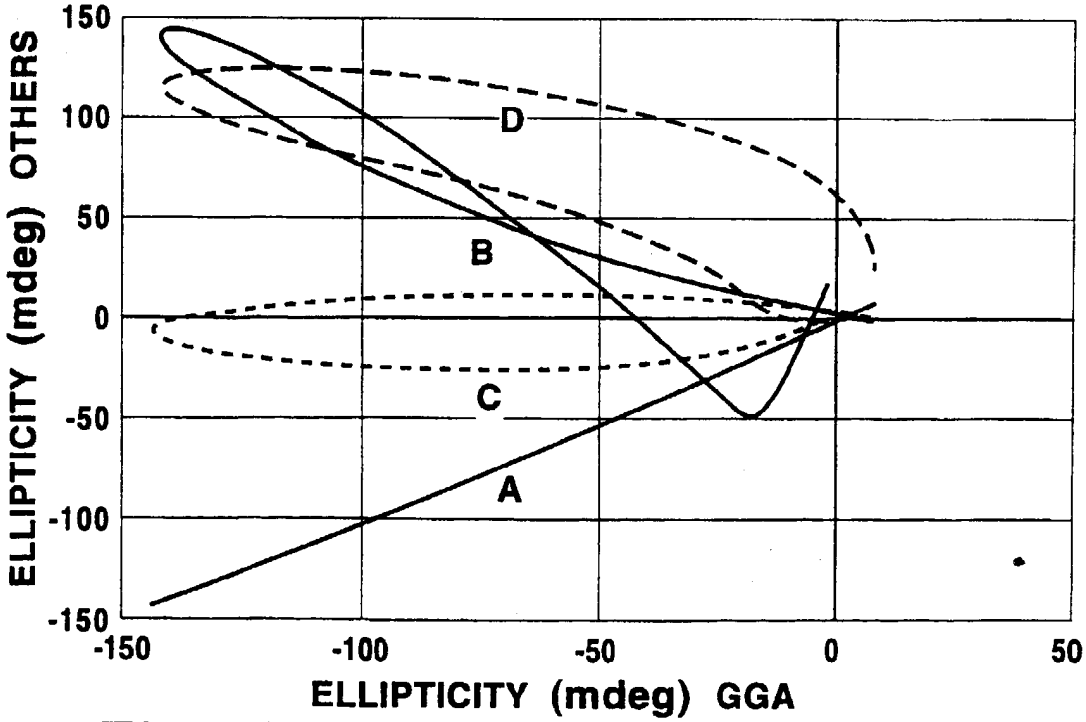
150
100
50
0
-50
-100
-150
ELLIPTICITY (mdeg) OTHERS
D
B
C
A
-150
-100
-50
0
50
ELLIPTICITY (mdeg) GGA

Fig. 14

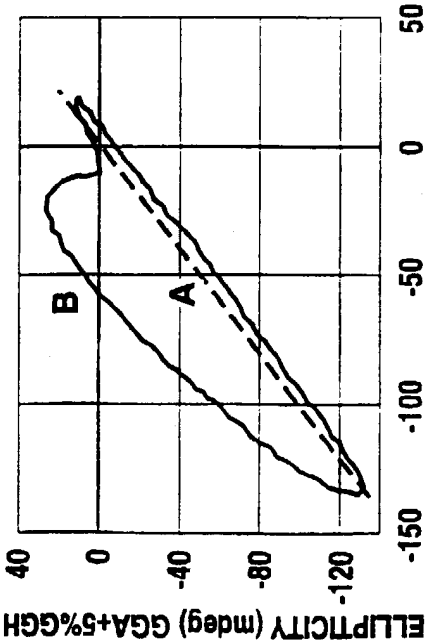

Fig. 15A

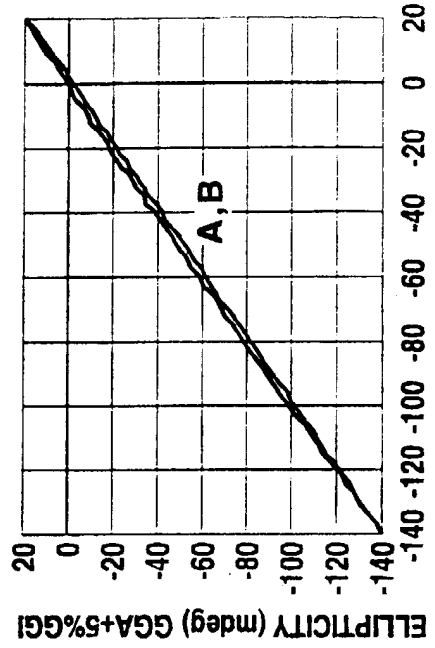

Fig. 15B

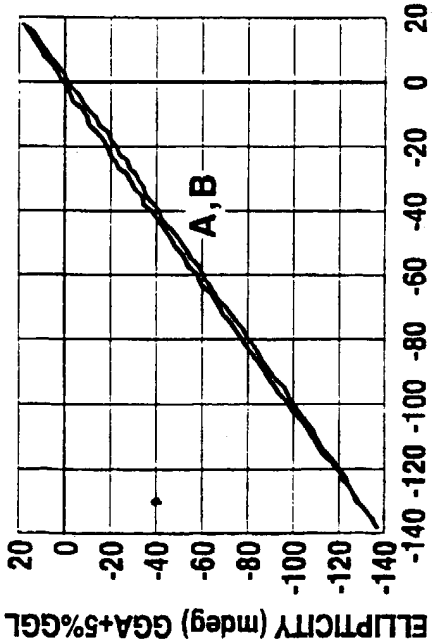

Fig. 15C

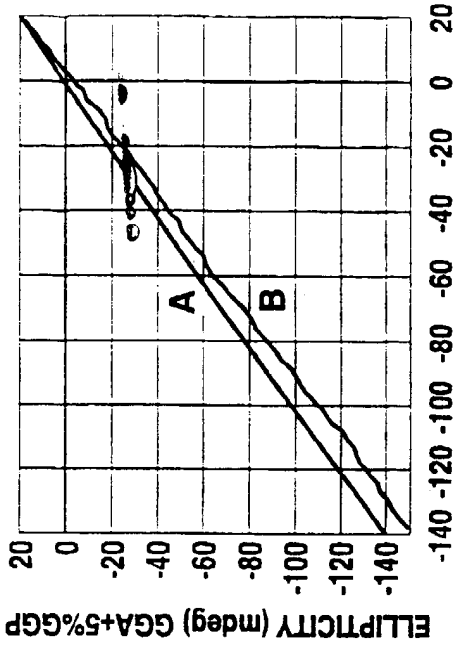

Fig. 15D

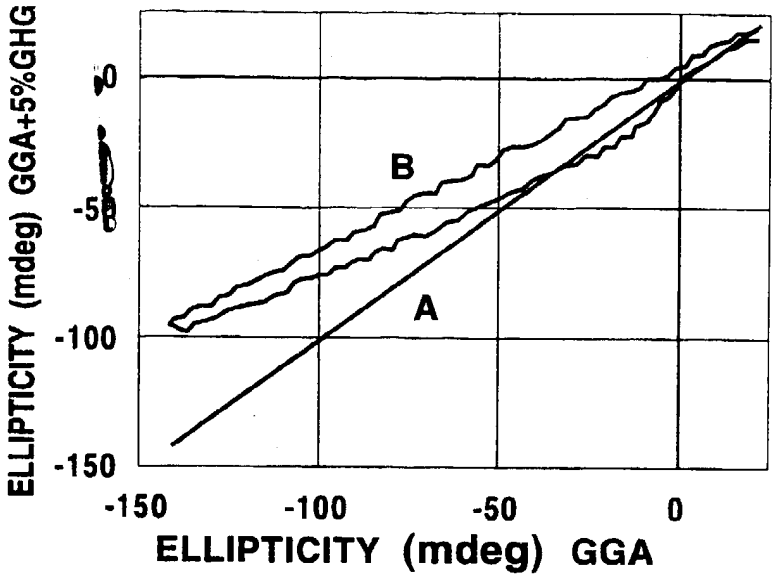

Fig. 16A

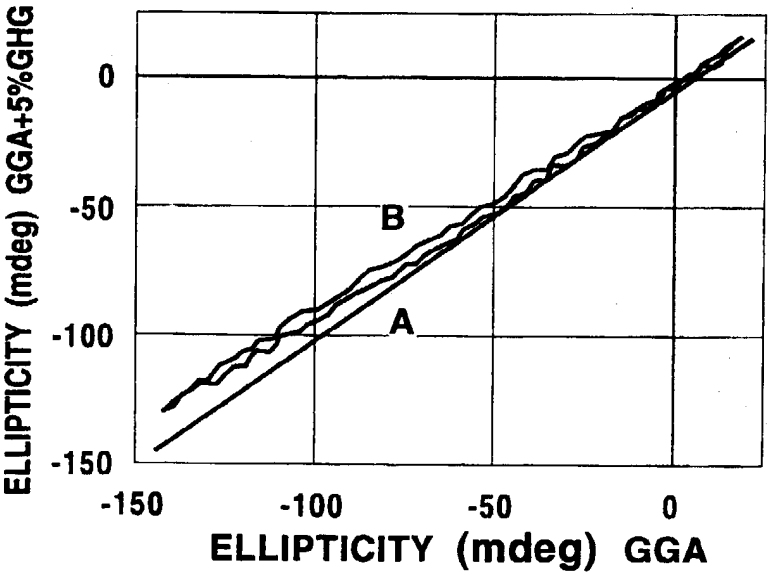

Fig. 16B

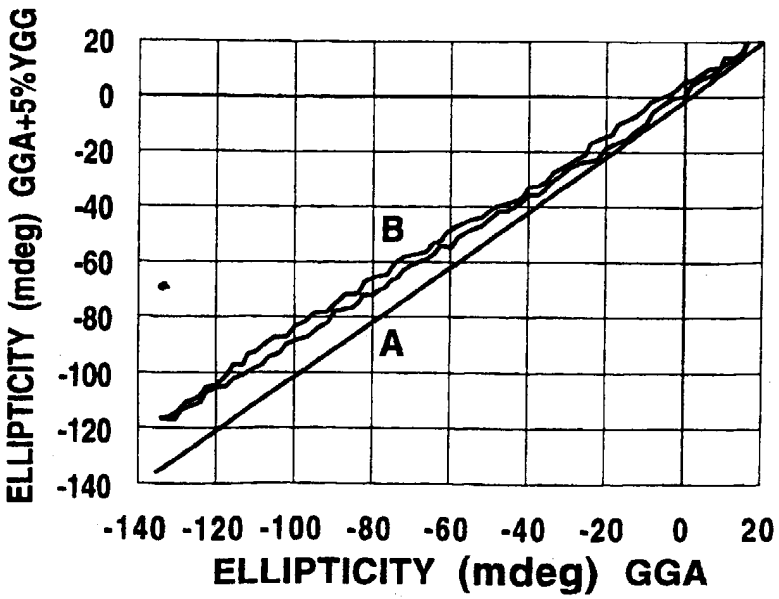

Fig. 16C

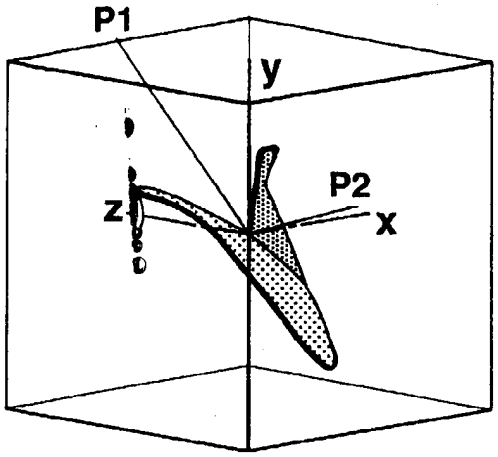

Fig. 17A

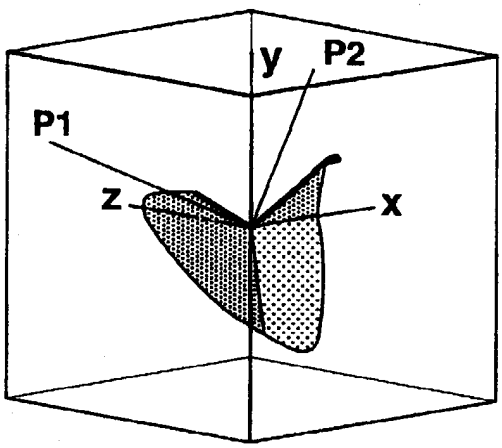

Fig. 17B

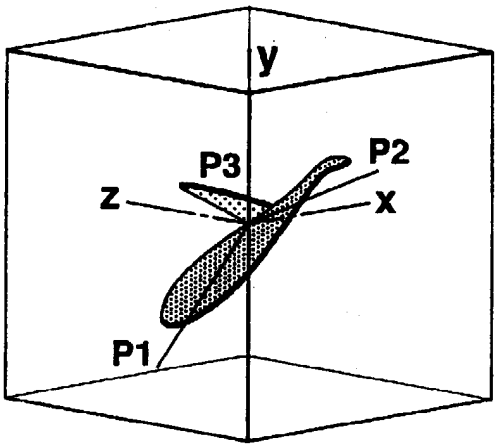

Fig. 17C

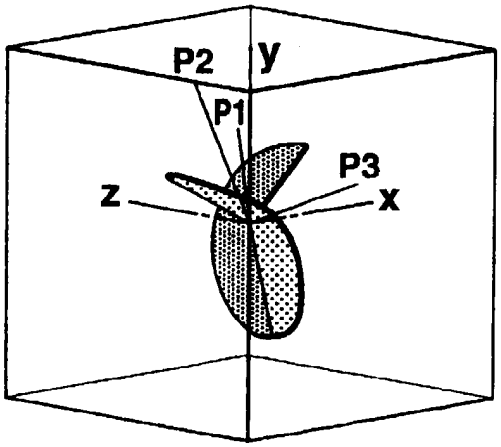

Fig. 17D

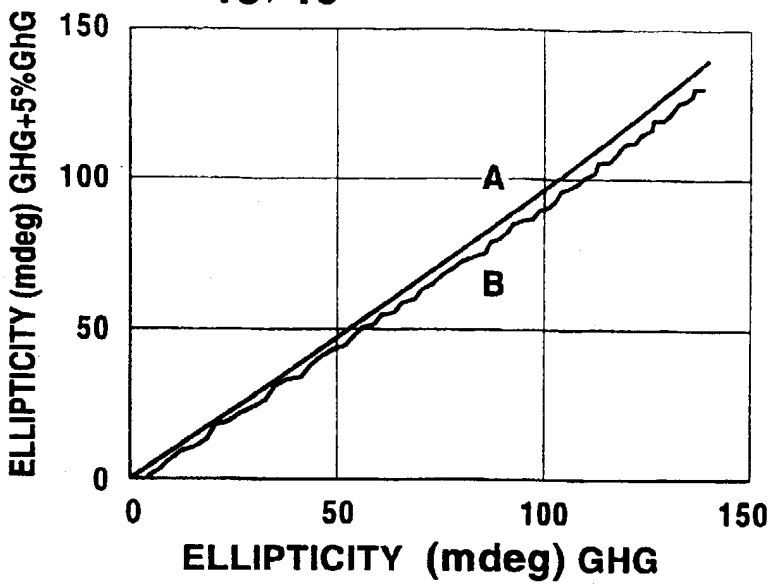

Fig. 18A

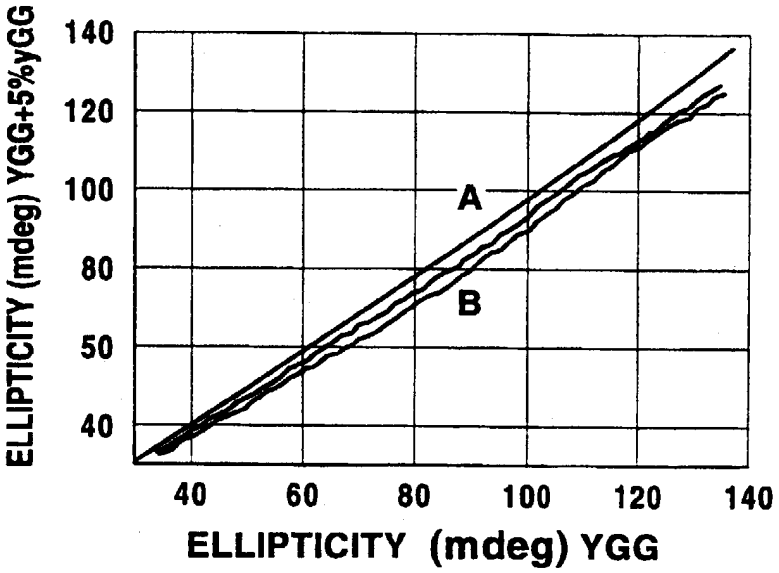

Fig. 18B

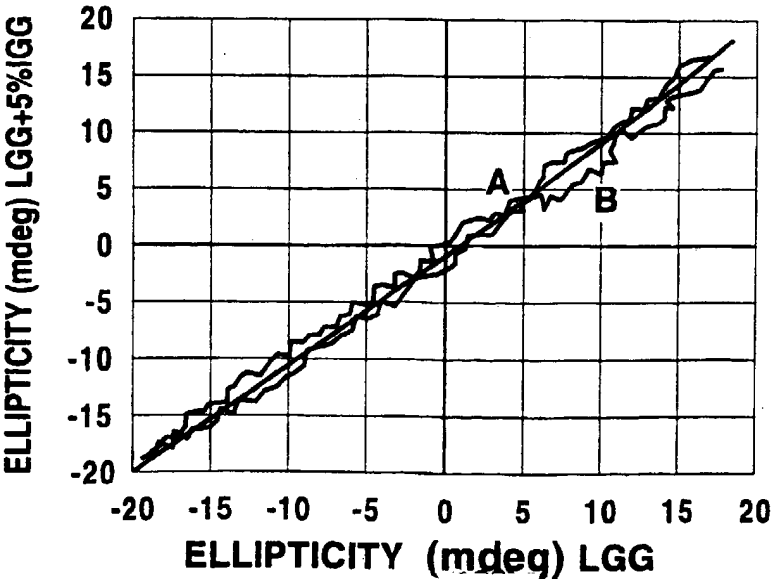

Fig. 18C

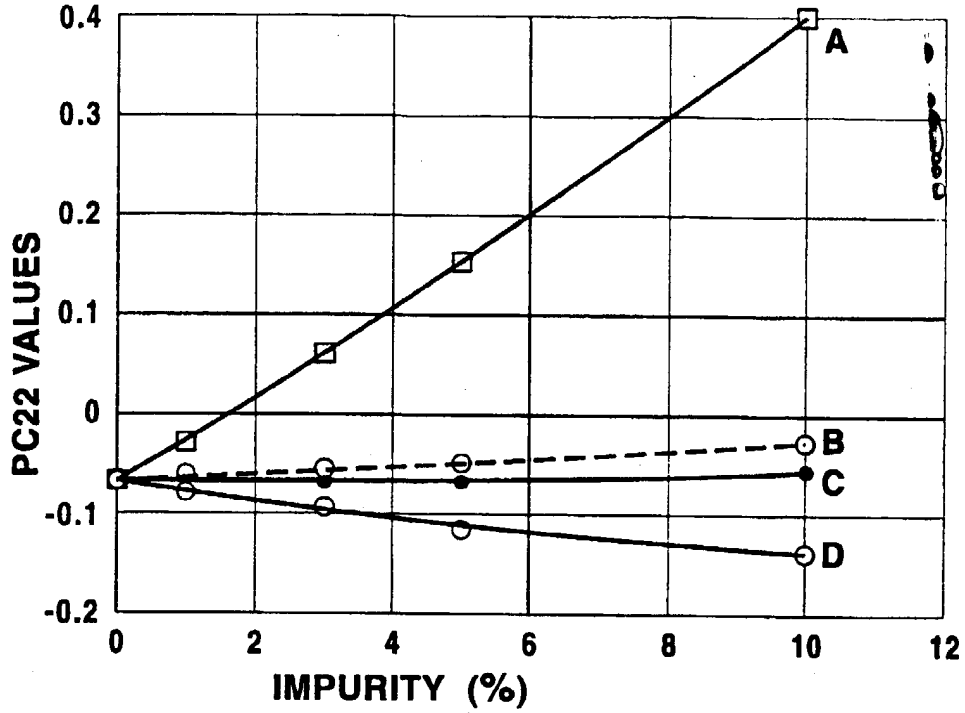
0.4
0.3
0.2
0.1
0
-0.1
-0.2
PC22 VALUES
0
2
4
6
8
10
12
IMPURITY (%)
A
B
C
D

Fig. 19

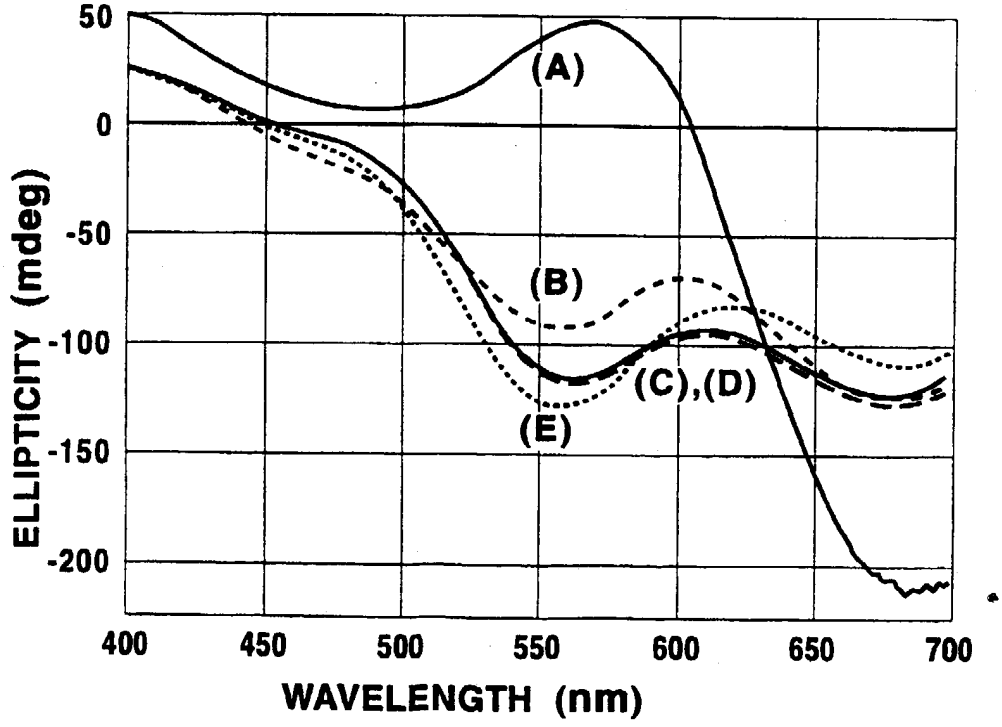
50
0
-50
-100
-150
-200
ELLIPTICITY (mdeg)
400
450
500
550
600
650
700
WAVELENGTH (nm)
(A)
(B)
(C),(D)
(E)

Fig. 20

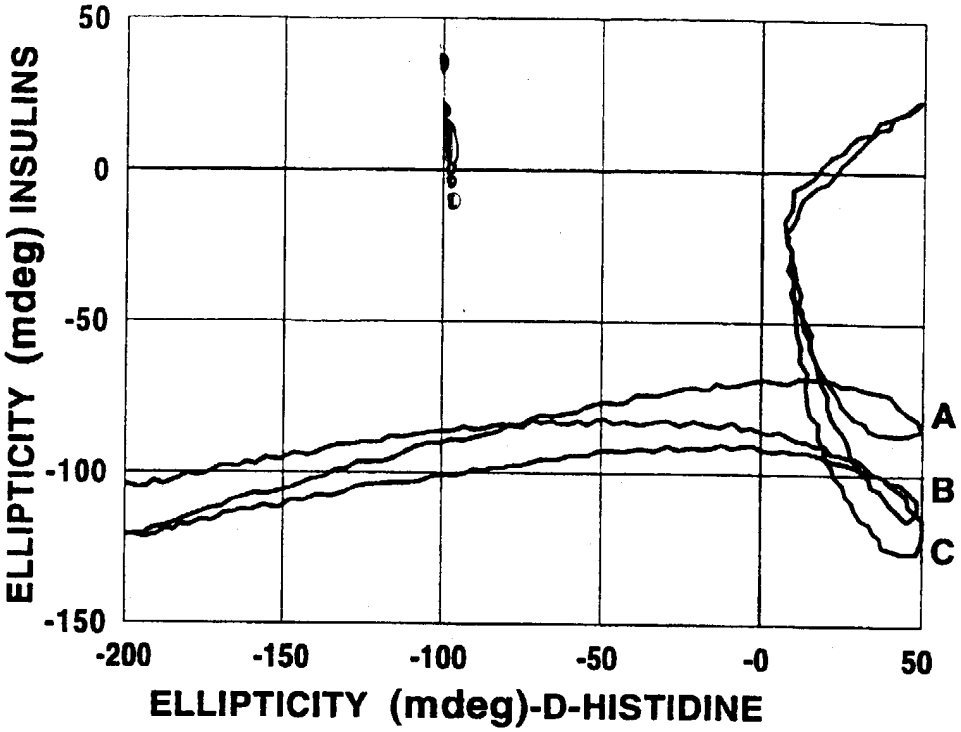
50
0
-50
-100
-150
ELLIPTICITY (mdeg) INSULINS
-200
-150
-100
-50
-0
50
ELLIPTICITY (mdeg)-D-HISTIDINE
A
B
C

Fig. 21

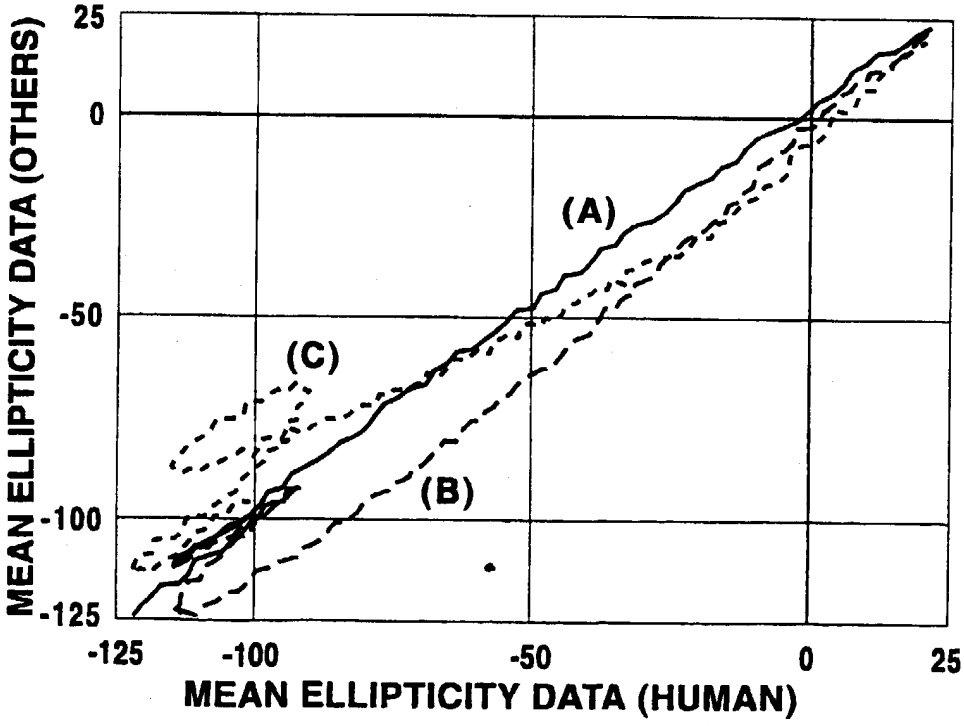
25
0
-50
-100
-125
MEAN ELLIPTICITY DATA (OTHERS)
-125
-100
-50
0
25
MEAN ELLIPTICITY DATA (HUMAN)
(A)
(B)
(C)

Fig. 22

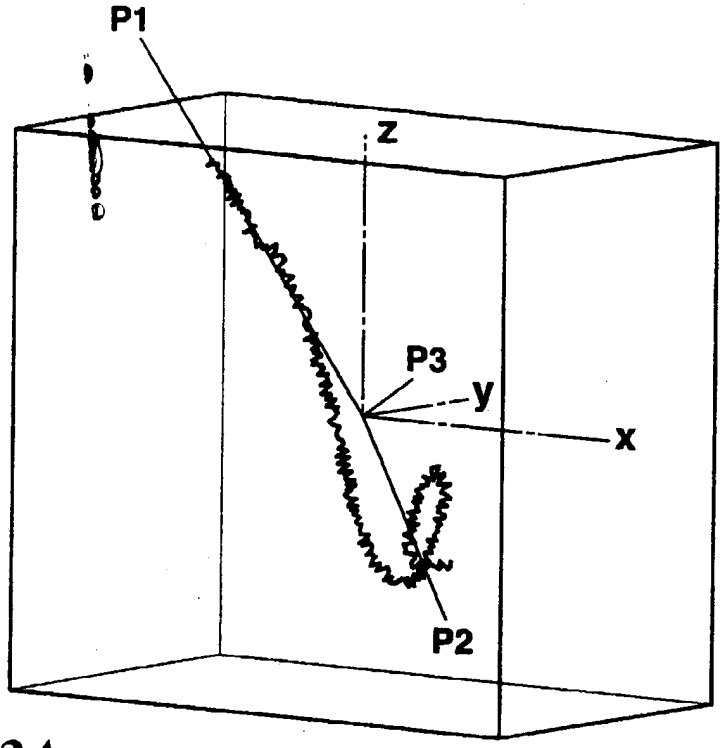

Fig. 23A

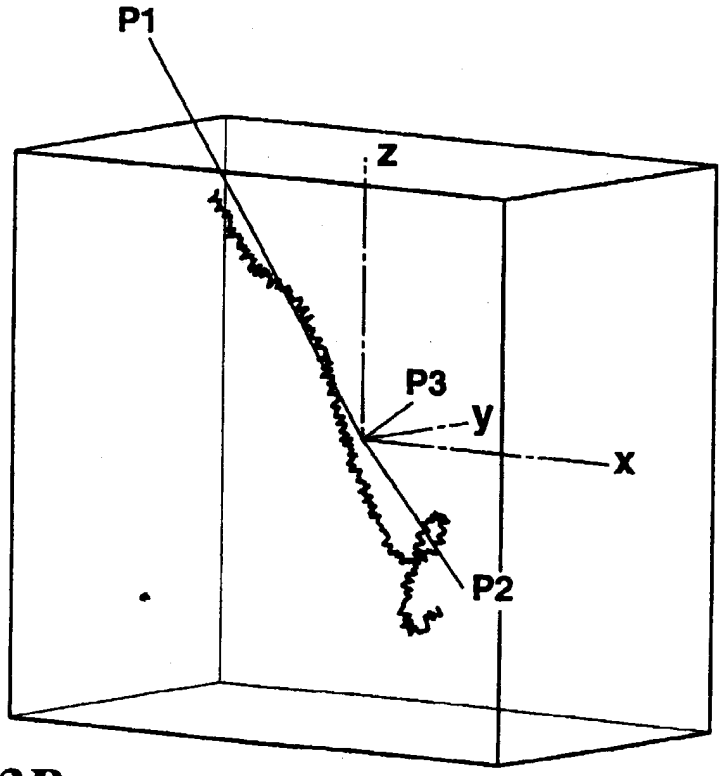

Fig. 23B